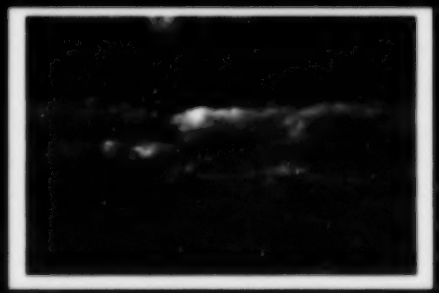
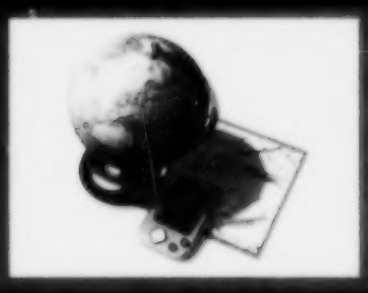




Evaluation of GPS Yield Mapping Technology at Reclaimed Industrial Sites in Alberta

August 15, 2006

ARECA

GeoTechnology Consultants

Prepared for:
Alberta Environment
Land Research User Group

Prepared by:
Ty Faechner, Ph.D., P.Ag.
Geotechnology Consultants
Dee Ann Benard, M.Sc., P.Ag.
Agricultural Research and Extension Council of Alberta
August 15, 2006

Pub. No. T/877
ISBN: 0-7785-4645-4 (printed edition)
ISBN: 0-7785-5060-5 (on-line edition)
Web Site: http://environment.gov.ab.ca

Any comments, questions, or suggestions regarding the content of this document may be directed to:

Environmental Monitoring and Evaluation Branch
Alberta Environment
Oxbridge Place
9820-106 St NW
Edmonton, AB T5K 2J6
Phone: (780) 427-6225
Fax: (780) 427-7958

Additional copies of this document may be obtained by contacting:

Information Centre
Alberta Environment
Main Floor,
Oxbridge Place
9820-106 St NW
Edmonton, AB T5K 2J6
Phone: (780) 944-0313
Fax: (780) 427-4407
Email: env.infocent@gov.ab.ca

TABLE OF CONTENTS

	Page
1.0 Introduction	1
1.1 Crop Yield Data	1
1.2 Objective	1
2.0 Methods	1
2.1 Analyzing Crop Yield Data	1
3.0 Results	4
3.1 Descriptive Analysis of Fields	4
3.2 Filtering Crop Yield Data	9
3.3 Descriptive Analysis of Industrial and Off-site Data	10
3.4 Spatial Variation of Crop Yield Data	10
4.0 Discussion	12
5.0 Applications	15
6.0 Conclusions	20
7.0 Recommendations	22
8.0 References	23
9.0 Appendices	26

LIST OF TABLES

Table 1 Crop yield measurements from 39 data sets in Alberta 4

Table 2 Field, industrial and off-site areas for the crop yield data sets in Alberta. 5

Table 3 Mean, variance and range of 13 unfiltered crop yield data sets from eight fields in Alberta selected for yield comparisons 6

Table 4 Unfiltered and filtered crop yield data for seven fields from different years in Alberta 9

Table 5 Comparison of GPS yield mapping to detailed site assessment for industrial sites in Alberta 13

Table 6 Means and confidence intervals for 39 data sets of crop yield measurements for industrial sites from eight fields in Alberta 14

Table 7 Mean, standard deviation, range and significance of 25 crop yield plots for the field 1 1997 well site 18

Table 8 Mean, standard deviation, range and significance of 25 crop yield plots for the field 1 1997 off-site 19

LIST OF FIGURES

Figure 1 Location of the clients with crop yield data 3

Figure 2 Field 1 crop yield spatial data for 1997, 1998, 2000 and 2003 7

Figure 3 Field 1, 2 and 3 boundaries, industrial site, off-site, and random off-site area 8

Figure 4 Field 4, 5 and 6 boundaries, industrial site, off-site, and random off-site areas 8

Figure 5 Field 7 and 8 boundaries, pipeline, off-site, and random off-site areas 8

Figure 6 Histograms for crop yield (bu/ac) from field 1 in 1997 16

Figure 7 Plot numbers for 20 X 20 m plots from the well site field 1 in 1997 17

Figure 8A and 8B Average crop yields of 20 X 20 m plots from the well site and off-site in field 1 1997 17

EXECUTIVE SUMMARY

Precision agriculture includes the use of crop harvesters equipped with yield monitors and moisture sensors to provide geo-referenced crop yield information. The objective of this study was to evaluate GPS grain yield mapping technology at reclaimed industrial sites in Alberta.

The Agricultural Research and Extension Council of Alberta (ARECA) and its member organizations contacted grain producers throughout Alberta with crop yield mapping capability. An interview protocol to acquire background information and crop yield data was completed by 47 farm operators. Six operators with appropriate crop yield data at reclaimed sites were selected for this study.

Thirteen crop yield data sets were generated for reclaimed industrial sites after removing erroneous crop yield observations. Spatial analysis of the crop yield data sets was used to produce a spatial average with confidence intervals for each industrial site and representative off-site area. The study explored the effectiveness of GPS yield monitoring technology at reclaimed industrial crop-land to:

- Provide a statistical comparison of site and off-site crop yield
- Determine the number of site and off-site data points available and necessary for a statistical comparison
- Determine field and site crop yield variability
- Evaluate crop yield at various shaped sites including linear (pipeline) and square or rectangular shaped sites
- Evaluate crop yield for one or more years from the same area
- Compare the value of adjacent or distant controls.

The study demonstrates that GPS yield monitoring technology is applicable for statistical analyses of crop yield on linear and non-linear industrial disturbances for one or more years. Data analysis requires expertise to filter and account for spatial variation. Statistical comparisons must be linked to field crop yield spatial variation. A 10 to 16 m sampling grid using GPS yield monitoring technology is required to account for spatial variation associated with the fields that have well sites. This means that for a typical 100 m by 100 m well site, at least 72 to 120 data points are required on-site and off-site to ensure a valid statistical comparison. The study also demonstrates that one or even several years of crop yield data may not be sufficient to provide assurance that site capability is acceptable under the influence of a full range of crop and environmental conditions typical for the area. Based on fields in this study, the area surrounding the industrial site can provide a satisfactory control for crop yield comparisons. The methodology developed in this study allows assessment of reclaimed sites using crop yield data from field scale equipment and determination of statistical confidence intervals.

ACKNOWLEDGEMENTS

We would like to recognize Alberta Environment's Land Research User's Group for financial support of this study. Thanks to the landowners who participated in this study and provided crop yield data for analysis. Thanks to the staff at Alberta Environment, Evaluation and Reporting Section, for their encouragement, and support of a unique approach for analyzing and evaluating crop yield data from landowners that have reclaimed land throughout Alberta. Thanks to Joy Vonk at ARECA for her work in completing the surveys.

1.0 Introduction

1.1 Crop Yield Data

Oil and gas wells in Alberta are being reclaimed and certified at a rate of approximately 2000 sites per year, which represents an estimated 2000 ha of land (Alberta Environment, 2005). This area does not include pipelines, gravel pits, coal mines, oil sands mines and plant sites. Many of these reclaimed sites are located on agricultural lands that are cropped annually with small grains, including wheat, barley, canola, flax and pulse crops. Current regulations administered by Alberta Environment require operators to reclaim and assess these sites visually and obtain a reclamation certificate. As part of the assessment process, operators at these reclaimed sites are required to report plant height and density and general health and vigor of the crop compared to the surrounding land, to meet the goal of equivalent land capability after reclamation. This qualitative assessment requires that crop growing on the reclaimed area meet or exceed 80% of the density and height of vegetation compared to a control area. Crop yield is the ultimate assessment of vegetation, and by inference, soil health. If these one-time visual crop estimates, along with several soil and landscape parameter estimates, meet the criteria, a certificate is issued to the operator (Alberta Environment, 1995).

Precision agriculture is a general term that includes the application of global positioning systems (GPS) and geographical information systems (GIS). Precision agriculture involves the use of harvest yield monitors and moisture sensors to provide many geo-referenced observations of crop yield per hectare at relatively low cost. This technology has prompted an interest in comparing crop yields from disturbed to non-disturbed sites. However, a serious challenge in making these comparisons is accounting for spatial correlation of the crop yield observations.

Field heterogeneity is systematically distributed throughout a field, with crop yields near one another more alike than crop yields farther apart (Fisher, 1931). Spatial correlation of crop yield has traditionally been addressed by reducing plot sizes so that the area is more homogeneous. Randomization and replication have also been used to minimize the effects of spatial autocorrelation. Classical statistics does not consider the effects of spatially correlated crop yield data. To account for spatial correlation in crop yield data, different spatial statistic techniques have been implemented (Dobermann and Ping, 2004; Griffin et al. 2006; Lauzon et al. 2005).

1.2 Objective

The main objective of this study was to determine the applicability of existing precision farming grain yield mapping technology to statistically assess crop yields at reclaimed industrial sites; and to evaluate the potential of grain yield mapping technology to identify and map crop yield variability and distribution at reclaimed industrial sites.

2.0 Methods

2.1 Analyzing Crop Yield Data

ARECA and its member organizations contacted grain producers throughout Alberta with crop yield mapping capability. For this project, it was essential that farm participants owned a combine harvester with a crop yield monitor and functional global positioning system. Combines equipped with these devices are able to record variables such as geographical

coordinates, grain flow rate, time, logging interval, distance traveled, swath width, grain moisture, and header status in intervals of one to several seconds. Farm operators with reclaimed industrial lands, including upstream oil and gas well sites, battery sites, plant sites, pipelines, sand and gravel pits, and road borrow sites, were considered for this study. An interview protocol (Appendix 2) to acquire background information about the selected sites was developed and completed by farm operators who participated in the study.

Farm operators who met the interview criteria were asked to submit historical crop yield monitoring data electronically for fields with reclaimed industrial sites. Crop yield data was initially assessed visually to ensure it was available for the entire field. If the crop yield map appeared satisfactory, farm operators were asked to provide crop yield monitor data for the submitted fields. Figure 1 illustrates the geographic locations of cooperators for this project. Geographic locations of industrial sites for this study were provided by each farm operator and confirmed by Alberta Environment. These sites were identified on crop yield maps or aerial photographs by farm operators.

Crop yield data can be erroneously measured and it is these measurements that need to be removed. Erroneous values can occur due to the combine header being down but before the grain flow has started or values are recorded (start-pass delay). Erroneous crop yield data can be collected at the end of a crop row when the header has not been raised (end-pass delay). Localized, extreme crop yield variation due to measurement error, crop management, or other events unrelated to the spatial patterns of crop yield variation can occur. The description of systematic and random errors in crop yield data is discussed by Griffin et al. 2005 and Simbahan et al. 2004. To remove these erroneous data, the raw crop yield monitor data in this study are filtered with Yield Editor 1.02 (Drummond, 2006) (Appendix 4, Figure A7) to provide thirteen data sets.

Selecting filtered crop yield data for each industrial site involved determining the industrial site's geographic location. Survey plans of the industrial sites provided by Alberta Environment were assigned a geographic coordinate system (latitude and longitude) using a raster image of each plan. Geographic locations were calculated based on distances and bearings from each industrial site. Filtered crop yield data were superimposed over each industrial site for the purpose of selecting crop yield data that occupied the same geographic space as the industrial site. This resulted in 39 data sets since an industrial, off-site, and random site was selected from each of 13 data sets.

On completion of selecting crop yield data for each industrial site, the 39 data sets were analyzed using geostatistical procedures to produce variograms for the industrial and off-site areas, using the public domain software GSLIB (Deutsch and Journel, 1998).

Sequential Gaussian simulation was used to generate data sets of crop yield using the mean, variogram and yield data from an area in a field. For each of the areas described in Table 1, data sets were created for the industrial site, off-site around the industrial site and randomly located off-site in the field. Crop yield from the access road into the industrial site was included in the off-site area. These data sets were used to produce a spatial average that is a mathematical mean computed using crop yield observations from the data set. Confidence intervals at the 95% level of significance were calculated from each data set. If the confidence intervals from an area did not overlap with those of another, crop yields of the two areas were significantly different.

Figure 1 Locations of the clients with crop yield data.

3.0 Results

3.1 Descriptive Analysis of Fields

Thirty-nine data sets of crop yield measurements were generated from eight fields with reclaimed industrial sites. The locations of the eight fields for crop yield data are displayed in Figure 1 and range from the Peace River area through central and southern Alberta. The industrial site and soil group for each of the fields included in the study is provided in Table 1. Most fields are in the Black Chernozemic soil great group with a variety of grain, oilseed and pulse crops grown on them. Three fields have historical crop yield data ranging from 2 to 6 years. There is mix of industrial sites from well sites to borrow pits to pipelines.

Table 1 Crop yield measurements from 39 data sets in Alberta.

Field	Year	Crop	Soil Group	Industrial Site	Owner
1	1997	Malt Barley	Black Chernozemic	Well site	A
1	1998	Malt Barley	Black Chernozemic	Well site	A
1	2000	Wheat	Black Chernozemic	Well site	A
1	2003	Peas	Black Chernozemic	Well site	A
2	2005	Peas	Dark Brown Chernozemic	Well site	B
3	2000	Peas	Dark Brown Chernozemic	Well site	C
3	2004	Wheat	Dark Brown Chernozemic	Well site	C
4	2002	Wheat	Dark Gray Chernozemic	Pipeline	D
5	2003	Barley	Black Chernozemic	Borrow Pit	E
5	2005	Canola	Black Chernozemic	Borrow Pit	E
6	2005	Wheat	Black Chernozemic	Pipeline	F
7	2004	Canola	Black Chernozemic	Pipeline	F
8	2004	Canola	Black Chernozemic	Pipeline	F

The areas of each field, industrial and off-site are provided in Table 2. The off-site refers to the area surrounding the industrial site or a randomly located site within the field. For comparison purposes, the area of the off-site was determined to be the same as the industrial site. Field sizes vary from 101 to 511 acres, while the industrial sites range from 2.1 to 4.3 acres.

Table 2 Field, industrial and off-site areas for the crop yield data sets in Alberta. Off-site refers to the area that is selected randomly in the field or surrounds the industrial site.

Field ID	Field (acres)	Industrial Site (acres)	Off-site (acres)
1	120	2.9	2.8
2	168	2.8	2.8
3	101	2.8	2.8
4	151	1.7	1.7
5	108	3.0	3.0
6	465	4.3	4.3
7	144	4.0	4.0
8	511	4.3	4.3

Crop yields calculated from raw yield data are provided in Table 3. The mean, variance, and range of field records have significant year to year variability. The mean of field 1 decreased by more than 40% from 1997 to 1998 while the range of yields declined by 25%. Field 7 and 8 had the same crop; however, the variance was much less for field 7, suggesting less variability in this field. Raw crop yield data for field 1 for years 1997, 1998, 2000 and 2003 had been filtered by the cooperator prior to this study. All other field data was not filtered prior to this study, and consequently contained erroneous crop yield data values. The high variance values and extreme range of crop yields for fields 2, 4, 5, 6 and 8 illustrate this effect of not being filtered.

Table 3 Mean, variance and range of 13 unfiltered crop yield data sets from eight fields in Alberta selected for yield comparisons.

Field	Year	Crop	Mean (bu/ac)	Variance	Range (bu/ac)
1	1997	Malt Barley	75.6	55.6	16 - 117
1	1998	Malt Barley	43.6	76.6	15 - 86
1	2000	Wheat	56.7	82.9	25 - 97
1	2003	Peas	40.7	31.2	16 - 60
2	2005	Peas	50.4	155236	0 - 53609
3	2000	Peas	43.6	312.2	5 - 379
3	2004	Wheat	55.9	211.1	5 - 208
4	2002	Wheat	21.0	14786	0 - 21770
5	2003	Barley	102.4	4841	0 - 4857
5	2005	Canola	98.3	279841	0 - 24177
6	2005	Wheat	54.1	69169	0 - 64215
7	2004	Canola	40.6	312.6	0 - 1792
8	2004	Canola	35.8	27889	0 - 255

Figure 2 shows crop yield data locations for field 1 for 1997, 1998, 2000 and 2003. Crop yield measurements were available at 5535 locations in 1998 compared to 6731 locations in 2000. This likely indicates that the combine harvester traveled at a faster speed in 1998 than in 2000. The remaining field boundaries, along with the industrial site, off-site and random site, are displayed in Figures 3, 4, and 5. Inset images of fields 3, 6, 7 and 8 are shown to allow better resolution of the industrial site or pipeline, off-site and random site (hence, one or more field boundaries is not evident). The choice of location for the random off-site areas was determined so that all locations within the field were equiprobable provided the same pattern could be duplicated (Deutsch and Journel, 1998). The pipelines for fields 4, 6, 7 and 8 transverse the field and are approximately 1550, 3160, 2895 and 3125 feet in length, respectively. Pipeline width at field 4 was 49 feet and at fields 6, 7 and 8 was 60 feet (communication with farm cooperators and Alberta Environment).

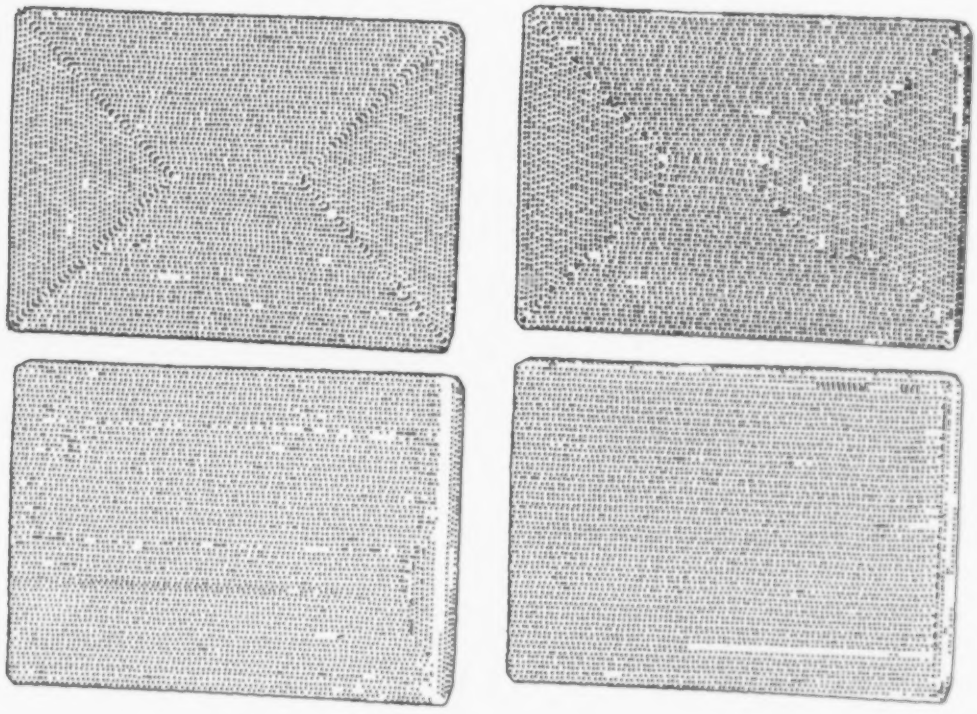


Figure 2 Field 1 crop yield data for 1997 (top, left), 1998 (top, right), 2000 (bottom, left) and 2003 (bottom, right). Yield data is measured at 5640, 5535, 6731 and 6588 locations for 1997, 1998, 2000 and 2003, respectively.

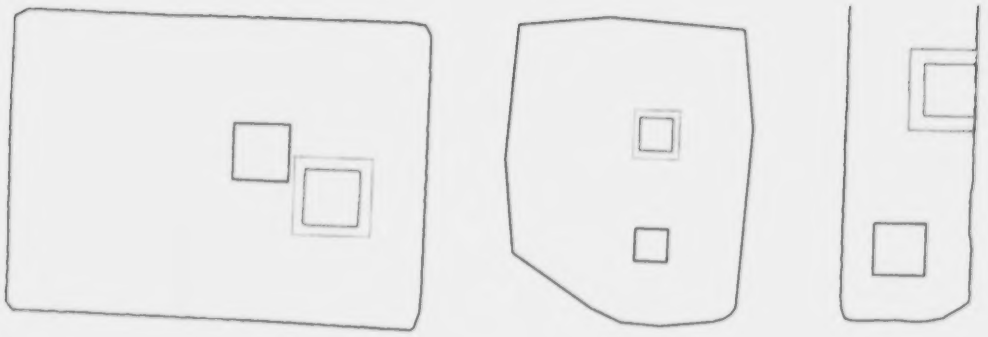


Figure 3 Fields 1, 2 and 3 boundaries (black), industrial site (red), off-site (green), and random off-site area (blue). These images are not to scale but illustrate the crop yield sampling areas for the industrial site, off-site and random sites.

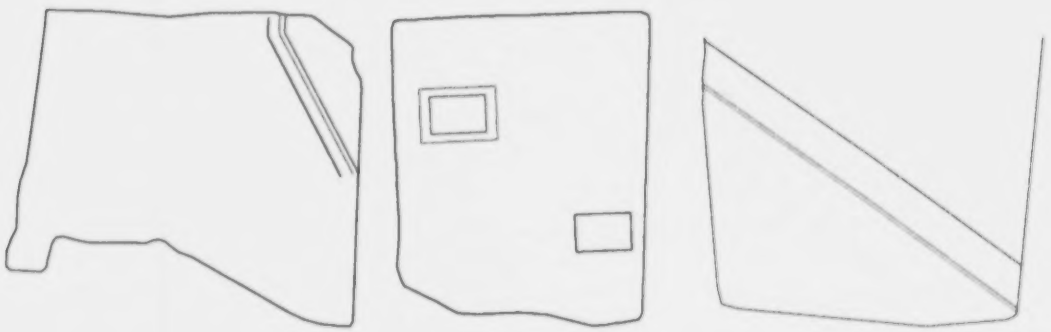


Figure 4 Fields 4, 5 and 6 boundaries (black), industrial site (red), off-site (green), and random off-site area (blue). The lines represent the pipeline on field 6 (red), off-site (green), and random off-site (blue). These images are not to scale but illustrate the crop yield sampling areas for the pipeline or industrial site, off-site and random sites.

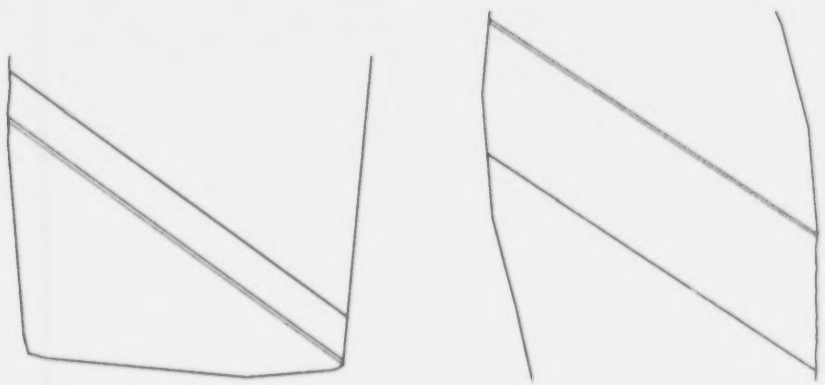


Figure 5 Fields 7 and 8 boundaries (black), pipeline (red), off-site (green), and random off-site area (blue).These images are not to scale but illustrate the crop yield sampling areas for the pipeline site, off-site and random sites.

3.2 Filtering Crop Yield Data

3.2.1 General Discussion

Filtering out erroneous crop yield measurements removed 7 to 22% of the measurements from analysis. Erroneous measurements are a function of the equipment, farm operator or environment. For example, in field 2 there were erroneous measurements of 0 and 53,609 bu/ac. When these measurements were removed, crop yield variability decreased, resulting in a more accurate representation of actual crop yield.

3.2.2 Technical Discussion

Filtering is the process of removing erroneous field harvest measurements of crop yield. Table 4 compares the mean, standard deviation (STD), coefficient of variation (CV), number of measurements, and range of the unfiltered and filtered crop yield for the images illustrated in Appendix 4, Figures A9 to A17. The variation is reduced while the CV is significantly lowered when comparing the filtered yield data to the unfiltered. The range of yield values changes as well and negative values and extreme outliers are removed from the data. There is usually about 15 to 20% less data as is evident in the reduction of the number of yield values. Field 2 had 38547 yield data values and with filtering this dropped to 32745, a 15% loss of data. The standard deviation diminished from 394 to 7.4 bu/ac and the CV decreased to a respectable 15.1%. The range narrowed from 0 - 53609 bu/ac to 21 – 74 bu/ac. Appendix 4, Figure A9 illustrates the raw yield data and the filtered yield data. One can observe the end passes and headlands where there are numerous, erroneous yield data.

Table 4 Unfiltered and filtered crop yield data for seven fields from different years in Alberta.

	Mean (bu/ac)	STD (bu/ac)	CV (%)	Number of Yield Values	Range (bu/ac)
Field 2 - 2005					
Filtered	49.4	7.4	15.1	32745	21 -74
Unfiltered	50.4	394.0	781.3	38547	0 - 53609
Field 3- 2000					
Filtered	44.0	13.1	29.7	8004	10 – 79
Unfiltered	43.6	17.7	40.5	10259	5 - 379
Field 3- 2004					
Filtered	56.3	12.8	22.7	16195	21 - 91
Unfiltered	56.0	14.5	26.0	18154	5 – 208
Field 4- 2002					
Filtered	20.2	2.6	13.0	30030	12 – 100
Unfiltered	21.0	121.6	580.1	32287	0 - 21770
Field 5- 2003					
Filtered	107.7	22.5	20.9	31150	34 – 176
Unfiltered	102.4	69.6	68.0	35356	0 - 4857
Field 5 - 2005					
Filtered	71.3	12.1	16.9	41635	30 - 110
Unfiltered	98.3	529.6	538.7	53363	0 – 24177
Field 6 - 2005					
Filtered	58.7	9.9	17.2	14877	24 - 90
Unfiltered	54.1	263.6	487.1	18480	0 - 64215

Field 7 – 2004					
Filtered	44.8	5.9	13.3	41480	22 - 66
Unfiltered	40.6	17.7	43.6	46839	0 – 1792
Field 8 - 2004					
Filtered	45.9	6.5	14.1	74954	10 - 79
Unfiltered	35.8	167.8	469.4	91688	0 - 255

3.3 Descriptive Analysis of Industrial and Off-site Data

3.3.1 General Discussion

Filtered crop yield data were selected for the well site, off-site and random site to provide 39 crop yield data sets. A descriptive, statistical analysis of each selected, filtered crop yield data set followed. Results from this analysis indicated that the means are different for the various industrial sites depending on the year, location and crop. Similarly, there is variation among the crop yield measurements within each data set; however, this variation is small and the data are clustered closely around the mean.

3.3.2 Technical Discussion

In the Appendix 1, the mean, variance, CV, skewness and kurtosis are provided for each field and industrial site. Three sites are assessed in each field including the well site (or borrow or pipeline), off-site and random site for all the fields and years in the study. The number of locations sampled for yield data range from 255 to 486 for fields 4 and 5, respectively. The coefficients of variation for the yield data are low and do not exceed 25%, except for the field 3 well site in 2000 that is 40%. The variance is reasonable, except for fields 3, 5 and 6, which have higher variability compared to the other fields. Similarly, yield data are positively and negatively skewed, with field 5 borrow pit, 2003 being the most positive at 1.7, and the random off-site for field 2, 2005 being the most negative at -1.21. The positive skewness indicates that the distributions are nearly symmetrical around the mean with the right tail of the distribution heavier than the left tail. A negative skewness means that the left tail of the distribution is heavier than the right tail. Most of the fields have crop yield values with a positive kurtosis, indicating a flatter distribution, except for the field 1, 1997 random and off-site areas, which are positive and very peaked.

3.4 Spatial Variation of Crop Yield Data

3.4.1 General Discussion

Crop yield measurements taken near each other tend to be more similar than measurements made farther apart. The distance over which these measurements are similar or correlated can be calculated. In this study of 39 crop yield data sets, crop yield measurements are similar or correlated over distances of 2 to 70 m (Appendix 4 - Table A1). Examination of historical data from field 1 indicates this variation is stable over time. On average, the distance over which yield measurements are similar or correlated for 27 data sets from fields 1, 2, 3, and 5 is 14 m. This means measurements sampled at greater distances are dissimilar or uncorrelated. Thus, to properly evaluate these data sets, crop yield measurements must be collected at distances of 14 m or less. Alberta Environment's current assessment with a 20 m grid exceeds this distance. In fact, 22 of 39 crop yield data sets required a finer resolution grid than 20 m.

Confidence intervals are an interval between two numbers where there is a specified level of probability that the numbers occur in this interval. To generate confidence intervals, a distribution of crop yield measurements with an average is needed for each industrial site. If the confidence intervals of crop yield measurements for two sites overlap, there is no statistically significant difference between the sites. Graphical and tabular displays of the confidence intervals illustrate these features. Confidence intervals show that in 15% of the 39 crop yield data sets, the yield in the industrial site was lower than the yield in the random off-site or off-site area.

Sampling in the area around an industrial site may be adequate to assess how successful reclamation is, provided there are a sufficient number of crop yield observations. There were 255 to 486 observations per crop yield data set in this study, with an average of 441 per industrial site (Appendix 1). One can obtain 550 to 600 observations per well site or control area with current GPS yield mapping technology. Currently, the Alberta Environment requirement is for 25 and 4 observations per well site and control area, respectively.

3.4.2 Technical Discussion

The spatial variation of crop yield at the eight fields in Alberta is presented in Appendix 4, Table A1. The experimental data for the well site in field 1, 1997 is modeled with a variogram having a nugget of 0.05 and a nested spherical structure with a contribution of 0.95 and range of 11 m for the structure. Similarly, the experimental data for the off-site in the same field is modeled with a spherical variogram that has a nugget of 0.02 and two nested structures with a contribution of 0.52 and range of 2.5 m for the first structure, and variance of 0.45 and range of 11 m for the second structure. For the random off-site variogram model, the nugget effect is 0.02, and there are two nested spherical structures with a variance of 0.61 at a range of 4.5 m and 0.37 at a range of 18 m.

Appendix 4, Figure A18 displays the model and experimental omni-directional variograms for each industrial site in field 1, 1997. The red dotted line represents the experimental variogram, while the dark dashed line represents the model variogram. There are no trends present in the experimental variograms of crop yield for each of the industrial sites. A trend is identified when the experimental variogram keeps increasing above the sill. This means that each crop yield variable is stationary, or its mean is independent of location and relevant to the entire field area.

The zone of influence or range of spatial correlation for crop yield varies among the industrial sites from 2 m for field 5, 2005, borrow pit to 70 m for field 8, 2004, off-site. Depending on the industrial site being evaluated, this indicates that there is a range of spatial variability. Further study is necessary to generalize these findings for more widespread implementation.

The range of spatial correlation for model isotropic variograms generally appears to be longest for pipelines, followed by well sites and borrow pits. There is only one borrow pit site included in this study, and caution is advised in extrapolating this result. The distance of correlation is 9 and 15 m in the borrow pit; however, outside this area, the spatial correlation is 2 to 6 m. Soil variability would appear to be large within fields and disturbing the soil when setting up an industrial site lessened this variability as determined by the range of spatial correlation. Length of time since these industrial sites have been reclaimed is unknown and this will impact the comparisons.

Environmental conditions such as moisture will affect crop yields. For the three cases where there is historical crop yield data (field 1, 3 and 5), the range of spatial correlation is very similar

temporally for each industrial site. For example, the well site in field 1 is spatially correlated at 11, 12, 15, and 14 m for 1997, 1998, 2000 and 2003. Higher variance values (Appendix 1) indicate more variability, which results in model variogram parameters having a higher variance contribution over longer distances. Comparing the 1998 and 2000 model variogram parameters for the off-site area in field 1, the variance contribution increases from 0.41 to 0.64 and the distance increases from 15 to 26 m. The variance increases from 36 to 59 for the off-site area over the same years (Appendix 1).

The range of spatial correlation averages 14 m for well and borrow sites for fields 1, 2, 3, and 5. This appears to be temporally stable. The current site assessment scheme (Alberta Environment, 1995) involves visually assessing plant height and density on a 20 X 20 m grid, such that there would be 25 observations collected for a typical well site of 100 by 100 m. A minimum of 4 locations is required outside the well site area for control sites. However, to account for the spatial variability of crop yield, 72 to 120 observations are required for the typical well site based on the 95% confidence intervals for these correlated distances. Additionally, another 72 to 120 observations are needed for comparison purposes in the surrounding area to address this spatial variability. The number of observations is determined by using a range of spatial correlation of 10 to 16 m and combine header width of 8 m. For a typical combine harvester traveling at 4 mph, crop yield measurements are recorded every second or 1.8 m. This means that 550 to 600 crop yield measurements are recorded by the combine harvester for a typical well site area.

The nugget is small for model variograms for well sites and larger for model variograms for pipelines. The nugget represents error in measurement or micro-scale variation too fine to detect, and as such, a well site has more clearly defined boundaries in the field compared to a pipeline. No consistent trend in the nugget effect is evident for borrow pits; this may reflect a temporal effect due to environmental conditions, since all three industrial sites had higher nuggets in 2005 compared to 2003.

4.0 Discussion

There are various methods for estimating or measuring crop yield. Machado et al. 2002 found that corn plant height explained 61% of the variation in grain yield in a study from Texas. A study by Girma et al. 2006 indicates that plant height, normalized difference vegetation index (NDVI), total nitrogen uptake, and chlorophyll content of winter wheat in a linear regression model had an R^2 value of 0.96 and coefficient of variation of 8.9%. As Girma et al. suggested the difficulty of utilizing plant height as a factor to better predict yield potential is that it must be coupled with some other parameter that takes into account the spatial variability of the plants. However, trying to use plant height and NDVI as reliable predictors of grain and oilseed yield may merit further investigation based on results obtained from these researchers. The correlation between visual estimates of plant height and crop yield has not been documented in western Canada.

Mid-season measurements of plants that are good predictors of grain yield and offer a window for sampling (Feekes 5 to 7) include NDVI, chlorophyll content, leaf color and plant height (Girma et al. 2006). Other plant characteristics, such as tillers per plant, are strongly associated with grain yield, but only at a specific growth stage, which narrows the prediction window. Correlation of plant density and height with grain yield needs to be evaluated to understand the strength of this relationship and how it may assist in designing a calibration tool for yield prediction.

Crop insurance organizations in western Canada utilize a formula to determine crop yield when visually inspecting crops prior to heading. This formula includes an estimate of the plant density, heads per plant, kernels per head and kernel weight on a square meter basis. Perhaps this may provide a methodology to integrate crop yield estimates derived from crop insurance organizations with plant height and density that is currently measured for reclamation assessment.

Grain yield mapping technology is assessed on reclaimed well site areas in this study. The mean, model variogram parameters and crop yield data from each industrial site are used to generate 101 unconditional, simulated data sets using sequential Gaussian simulation. The data sets are spatially averaged and those averages form the basis for determining an industrial site's confidence intervals at the 95% level of significance.

Grain yield mapping technology that is evaluated in this study can provide an objective, statistically sound measure to assess well site reclamation efforts. The current detailed site assessment criteria (Alberta Environment, 1995) rely on visual assessment, which is a qualitative, subjective measure that does not account for spatial variation. A comparison of yield mapping using GPS and detailed site assessment is provided in Table 5.

Table 5 Comparison of GPS Yield Mapping to Detailed Site Assessment for industrial sites in Alberta.

	Crop Yield Assessment Method	
	GPS Yield Mapping	**Detailed Site Assessment as per Criteria***
Control Observations Required	72 to 120	4
Well Site Observations Required	72 to 120	25
Control Observations Available	550 to 600	-
Well Site Observations Available	550 to 600	-
Control Data Accuracy (error)	1 - 3%	Unknown
Well Site Data Accuracy (error)	1 - 3%	Unknown
Confidence Intervals at Well Sites	95% or 99%	None
Confidence Intervals at Controls	95% or 99%	None

*Alberta Environment, 1995

Two controls are considered in this study: one immediately surrounding the industrial site, called the off-site, and one randomly located in the field. The off-site area represents what Alberta Environment advises lease owners to consider as a control when conducting an assessment of the site. In some cases (15%), yields are considerably lower in the industrial site (Table 6), than in either of the off-sites.

It is best to choose a control site (off-site) near the reclaimed industrial site when making yield comparisons to determine if equivalent land capability has been achieved in the reclaimed industrial site. In most cases, there is a consistent trend showing yields from the random site and off-site areas being both lower than yield from the industrial site or vice versa. However, yield data from a nearby off-site is likely to be more consistent. Average yield for the whole field should not be used, as it is less reliable than choosing an area immediately around the industrial site. In this study, mean crop yields for the off-site area are statistically significantly higher in half of the cases compared to the average crop yield for the field (Tables 6 and 4). Use of the field average tends to either underestimate or overestimate the variation in yield. For example,

the field average is 56 bu/ac for field 3 (2004) compared to the values of 58 and 61 for the off-site and well site, respectively.

Table 6 Means and confidence intervals for 39 data sets of crop yield measurements for industrial sites from eight fields in Alberta.

Industrial Site	Mean (bu/ac)	Confidence Interval* (bu/ac)	
Field 1 - 1997			
Well site	61.3	61.2	61.4
Off-site	73.6	73.6	73.7
Random Off-site	75.9	75.8	75.9
Field 1 - 1998			
Well site	48.2	48.0	48.2
Off-site	44.4	44.4	44.5
Random Off-site	38.4	38.4	38.3
Field 1 - 2000			
Well site	53.2	53.1	53.2
Off-site	53.7	53.6	53.7
Random Off-site	50.2	50.1	50.2
Field 1 - 2003			
Well site	39.8	39.7	39.8
Off-site	40.5	40.5	40.5
Random Off-site	33.9	33.8	33.9
Field 2 - 2005			
Well site	52.4	52.2	52.4
Off-site	52.7	52.6	52.8
Random Off-site	46.2	46.0	46.2
Field 3 - 2000			
Well site	38.0	37.6	38.1
Off-site	43.3	43.2	43.7
Random Off-site	47.3	47.2	47.4
Field 3 - 2004			
Well site	59.0	58.9	59.1
Off-site	57.8	57.8	58.0
Random Off-site	60.8	60.4	60.9
Field 4 – 2002			
Pipeline	16.5	16.5	16.5
Off-site	16.4	16.4	16.5
Random Off-site	17.0	16.9	17.0
Field 5 – 2003			
Borrow Pit	65.6	65.6	65.8
Off-site	91.4	91.3	91.6
Random Off-site	111.6	111.6	111.9
Field 5 – 2005			
Borrow Pit	71.2	71.1	71.3
Off-site	75.1	74.9	75.2
Random Off-site	72.4	72.3	72.5

Field 6 – 2005			
Pipeline	60.2	60.0	60.5
Off-site	59.5	59.3	59.7
Random Off-site	55.5	55.3	55.7
Field 7 – 2004			
Pipeline	47.1	47.0	47.2
Off-site	45.7	45.6	45.8
Random Off-site	45.3	45.3	45.4
Field 8 – 2004			
Pipeline	50.9	50.9	51.0
Off-site	47.9	47.9	48.0
Random Off-site	44.6	44.6	44.7

*95% level of significance

Control sampling in the immediate vicinity of an industrial site should be sufficient to evaluate the effectiveness of the reclamation effort, provided an adequate number of observations are made. In this study, the minimum number of observations for the random or off-site areas is 255, with an average of about 441 per area. To detect a 5 bu/ac difference with a standard deviation of 7.6 bu/ac, a minimum of 57 observations is required for a statistical comparison at the 5% level of significance. In contrast, the 20 X 20 grid recommended by Alberta Environment would provide only 25 observations for the industrial site, and only four for the off-site or random area.

5.0 Applications

Crop yield data can be used to determine similar and different areas within the well site and off-site, as well as location, extent and degree of yield limitations where yield is significantly different than off-site. Histograms and grids of crop yield can be used to demonstrate these yield distributions and differences (Figures 6, 8A and 8B).

5.1 Histograms

Histograms of yield data from the well site and off-site or control area can provide a graphical comparison of frequency distributions of crop yield. If these distributions are different, the extent and location of these low yielding areas on the well site can be identified. For example, histograms of 1997 barley yield for the well site and control area in field 1 are displayed in Figure 6. The two distributions are almost non-overlapping. If the criterion for reclamation is based on crop yield from the well site, and it must be greater or similar to the minimum crop yield from the off-site area, then 96% of the crop yield measurements from the well site area met this criterion. However, 47% of these measurements from the well site are less than 60 bu/ac, which is the third lowest crop yield value from the control area. Consideration must be given to the application of this methodology for identifying the area and amount of crop that does not meet the goal of equivalent land capability.

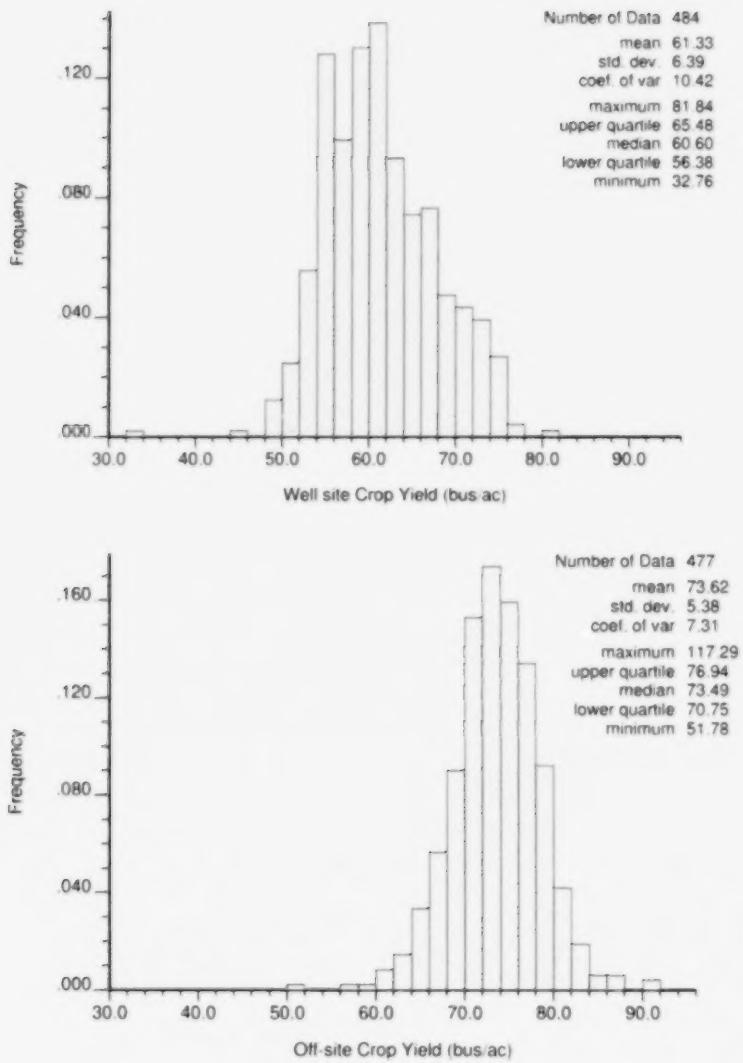


Figure 6 Histograms (top – well site and bottom – off-site) for crop yield (bu/ac) from field 1 in 1997.

5.1 Grid Analysis

Crop yield differences and similarities can be demonstrated by comparing and grouping crop yield means from each 20 x 20 m grid as well as the 15 to 20 individual yield points within each grid. For example, the average yield of 54.5 bu/ac shown in Figure 8A for well site grid #13 can be statistically compared with yields from other grids, on and off-site. In addition, yield variability within grid #13 can be determined (5.1 bu/ac range) and compared with other grids as shown in Table 7.

Statistical comparisons can be easily determined within and between grids since approximately 15 to 20 yield data points are available from each grid. In this study, average

	6	7	8	9	10	
5	21	22	23	24	25	11
4	16	17	18	19	20	12
3	11	12	13	14	15	13
2	6	7	8	9	10	14
1	1	2	3	4	5	15
	20	19	18	17	16	

Figure 7 Grid numbers for 20 X 20 m grids from the well site and off-site area in field 1 1997 where average crop yields (bu/ac) were determined for each grid. Grid numbers refer to each grid shown in Tables 7 and 8.

66.4	68.3	63.7	65.7	72.4
59.3	62.9	59.4	63.3	64.6
54.2	56.3	54.5	60.3	61.4
56.4	53.8	55.6	61.0	60.1
66.3	57.7	58.6	68.7	68.5

A

	73.6	73.6	71.2	72.0	77.9	
75.1	66.4	68.3	63.7	65.7	72.4	77.6
71.1	59.3				64.6	78.6
74.0	54.2				61.4	73.0
64.1	56.4				60.1	66.0
76.4	66.3	57.7	58.6	68.7	68.5	72.7
	72.8	68.6	66.3	74.4	73.8	

B

Figures 8A and 8B Figure 8A shows average crop yields (bu/ac) at twenty five 20 X 20 m grids at Field 1, 1997. Each color represents grids with crop yield means that are statistically similar at P=0.05 using analysis of variance and Tukey's test. Figure 8B shows 16 on-site (yellow) and 25 off-site (green) grid yields where on-site/off-site crop yields are significantly different at P=0.01 using a Student's t-test.

crop yields were determined at 25 numbered grids within and 20 numbered grids outside the well site area in field 1 1997 (Figures 7, 8A and 8B; Tables 7 and 8). Four separate yield comparisons were made including:

- **comparison of mean grid crop yields within the well site** - to identify crop variability as well as its location and extent (Figure 8A, Table 7). This comparison shows that there are 4 significantly different (one-way analysis of variance P=0.05) yield patches in the well site of 53.8 – 54.5, 61.0 – 61.4, 65.7 – 68.7 and 72.4 bu/ac with a difference of 18 bu/ac or 25% between the highest (grid 25) and lowest mean yields (grid 11). Locations of these significantly different yield patches are along the north and south lease edges for the highest yields while the lowest yields are in the center-west portion of the well site. The location of various yield patches identifies areas that may require further remediation of the soil such as the center-west portion of the site. This yield comparison can facilitate site assessments and mitigation planning to address soil disturbances.
- **comparison of 16 well site with 20 off-site grid crop yields** - to determine differences between well site and off-site grids (Figure 8B). This comparison shows a significant yield difference between well site (61.3 bu/ac) and off-site (73.6 bu/ac) grids (Student's t-test P=0.01), a 17% difference. It also indicates that areas of lowest well site yield (yellow grids) are confined to the well site, and therefore by implication, related to well-site impacts rather than background conditions.
- **comparison of equivalent areas** – to determine the extent of well site area where crop yield is equivalent to off-site yield. This method of determining equivalency is similar to that used for assessing large reclaimed areas such as surface mines where the expectation is that the reclaimed mine landscape will contain the same areal distribution of land capability classes as pre-mine conditions. For example, if the pre-mine landscape had a land capability class distribution of Class 1 – 20%, Class 2 – 40%, and Class 3 – 40%, the expectation would be that the reclaimed landscape would contain the same distribution of land capability classes. When this method is applied to Field 1 (barley, 1997), approximately 6 of 25 (24%) well site grid yields are similar to controls and 19 of 25 (75%) are not. In 2003 (peas), 10 of 25 (40%) grid yields are equivalent to controls and 15 (60%) are not. Non-equivalent grid area locations coincide for both crops indicating the presence of limiting soil or site conditions in the central-western portion.
- **comparison of within-grid variability at 25 on and 20 off-site grids (Tables 7 and 8)**. There were 16 to 20 yield data points available at each grid to determine grid means. This comparison shows that data point yields can vary widely within grids. For example, well site grid #2 had a range of 35.2 bu/ac (highest yield minus lowest yield) and off-site grid #2 has a yield range of 80.8 bu/ac. This clearly indicates that many yield assessment points are required in each grid to make an accurate determination of variability and establish statistical differences. Accurate site and off-site yields cannot be assured with one or two measured or visual assessment points per grid.

Table 7 Crop yield mean, standard deviation, range and significance at 25 grids for the Field 1 1997 well site.

Plot	Mean (bu/ac)	STD (bu/ac)	Range (bu/ac)	Significance*
25	72.4	1.8	6.3	a
4	68.7	5.2	14.3	b
5	68.5	3.8	12.3	bc
22	68.3	3.5	11.9	bc
21	66.4	3.4	13.0	bcd
1	66.3	4.1	14.8	bcd
24	65.7	6.2	17.2	bcd
20	64.6	3.9	12.1	de
23	63.7	2.4	9.0	def
19	63.3	4.4	13.7	defg
17	62.9	3.8	13.6	defg
15	61.4	2.1	6.5	efgh
9	61.0	2.6	10.2	efgh
14	60.3	1.7	5.3	fghi
10	60.1	2.5	8.1	fghi
18	59.4	3.0	9.0	ghij
16	59.3	4.6	14.7	ghij
3	58.6	2.0	7.7	hij
2	57.7	8.0	35.2	ijk
6	56.4	2.4	9.4	ijk
12	56.3	1.6	6.4	ijk
8	55.6	1.1	4.2	jk
13	54.5	1.6	5.1	k
11	54.2	3.9	13.3	k
7	53.8	1.7	6.0	k

*Crop means differing by letter are significantly different at P=0.05 using Tukey's test.

Table 8 Crop yield mean, standard deviation, range and significance of 20 crop yield grids for the Field 1 1997 off-site.

Plot	Mean (bu/ac)	STD (bu/ac)	Range (bu/ac)	Significance*
12	78.6	4.3	14.3	a
10	77.9	2.2	7.4	ab
11	77.6	1.7	6.5	ab
1	76.4	5.1	18.6	ab
5	75.1	2.4	7.5	abc
17	74.4	6.1	18.1	abc
3	74.0	6.9	20.8	abc
16	73.8	3.0	11.3	abc
7	73.6	2.5	10.2	abc
6	73.6	3.4	9.4	abc
13	73.0	5.4	15.2	abc
20	72.8	1.9	8.5	abcd
15	72.7	1.8	7.4	abcd
9	72.0	2.8	8.8	abcde
8	71.1	1.4	4.2	bcde
4	71.1	1.5	5.8	bcde
19	68.6	3.5	14.9	cdef
18	66.2	3.2	10.8	def
14	66.0	3.1	13.0	ef
2	64.1	20.2	80.8	f

*Crop means differing by letter are significantly different at P=0.05 using Tukey's test.

6.0 Conclusions

This study proves that GPS crop yield monitoring technology can provide a precise, practical and quantitative methodology for assessing crop yield at reclaimed industrial sites in Alberta. This technology has features that offer a number of advantages over current crop assessment methods used at reclaimed industrial sites. These features include:

- **Proven technology** – GPS crop yield monitoring technology has been tested and proven throughout the world with conventional grain harvesting equipment and is available in Alberta. Because of its demonstrated use for assessing crop yield, this technology has a high degree of acceptance among agricultural stakeholders.
- **Rapid and objective measurements** – once crop yield monitoring equipment is installed on harvest machines, crop yield data are generated objectively and without any additional effort during harvest operations. Quantitative data can be generated each year at the same locations to confirm yield patterns or provide additional information about year to year variability associated with different crops and moisture conditions.

- **Accurate/reliable measurements** – numerous crop yield data can be generated with a high degree of accuracy (1 to 3% error) after filtering. The large number of data within an area affords opportunities for statistical analyses. In this study, 3274 filtered crop yield measurements were obtained from 25 acres of industrial site area for an average of 127 measurements per acre. This compares to 9 measurements/acre taken at a typical well site under current criteria (Alberta Environment, 1995). An average of 32 to 48 measurements/acre is necessary for statistical analyses. Statistical yield comparisons can provide accurate information about on-site/off-site equivalency and on-site variability. The on-site yield variability information can also be used as a reliable indicator of location, extent and degree of on-site soil variability.
- **Standard method** - GPS crop yield monitoring technology can be applied to all harvested seed crops at industrial sites with the same level of confidence.
- **Monitoring** - this technology can be applied routinely on the same sites year after year as a tool to track temporal changes in crop yield associated with crop varieties, management, and environmental conditions.
- **Variability** – crop yield variability can be accurately mapped at a high level of detail on linear and non-linear industrial sites. The range and mean of crop yields within a well site illustrated the large, significant variability that occurred over short distances.

The key finding of this study is the importance of crop yield variability and the need to obtain sufficient crop yield measurements to perform statistical comparisons of industrial sites with off-site controls. Analysis of crop yield from sites used in this study show that a minimum of 72 to 120 crop yield measurements are required on-site and off-site to provide statistical comparisons and answer the questions "are on-site and off-site yields similar or different?" and "how different are they?".

Results from this study show that the number of crop yield measurements for statistical analyses considerably exceeds those currently used for on-site/off-site comparisons at well sites (Alberta Environment, 1995). This study demonstrates that GPS yield monitoring technology is a tool that can be used to accurately and objectively answer the question "has equivalent land capability been achieved on an industrial site"?

7.0 Recommendations

A precise, practical and quantitative methodology for assessing crop yield at reclaimed industrial sites in Alberta has been developed. To evaluate its potential application over a broad range of conditions, it is recommended that a more intensive study be conducted in the following areas:

- **Range of industrial sites** – explore the use of this technology to include the full range of reclaimed industrial sites (all categories and ages of well sites, pipelines, batteries, borrow sites, sand and gravel pits, coal mines and rail lines). In addition, the number of industrial sites assessed in each category should be increased to improve confidence in the methodology and conclusions developed from this limited study.
- **Crop diversity** – assess crop yield data from a variety of crops such as oats, rye, triticale, and forages for seed.
- **Temporal response** - assess crop yield annually at a variety of sites for a number of years to determine crop yield response to management and environmental variables.
- **Geographic differences** – assess crop yield in all ecoregions of the white zone to determine response to a variety of soil, environmental and management conditions.
- **Availability of GPS technology** - assess the availability and expertise of GPS crop yield monitoring technology in Alberta for the purpose of applying it to crop yield assessments at reclaimed industrial sites.
- **Technology improvement** - the nature of this technology lends itself to further data analyses and discoveries at reclaimed industrial sites. Information potentially available from further analysis of this technology includes:
    - Crop yield variability attributed to site reclamation versus field variability,
    - Correlation of crop yield variability with site and soil variability,
    - Correlation of crop yield variability with site and soil properties,
    - Correlation of GPS crop yield measurements with visual crop assessment parameters.

Assessing crop yield data collected with GPS crop yield monitoring technology in these areas will assist in identifying the availability and importance of the technology.

8.0 References

Alberta Environment. 1995. Reclamation Criteria for Wellsites and Associated Facilities-1995 Update. Available: http://www3.gov.ab.ca/env/protenf/landrec/documents/wellsite_criteria_1995.pdf (Accessed March 1, 2006).

Alberta Environment. 2005. State of Environment Reporting – Land. Available: http://www3.gov.ab.ca/env/soe/land_indicators/40_oilgas_reclamation.html (Accessed March 1, 2006).

Burks, T. F., Shearer, S. A., Fulton, J. P. and C. J. Sobolik. 2003. Combine yield monitor test facility development and initial monitoring test. Applied Engineering in Agriculture 19:5-12.

Bourgault, G., Journel, A. G., Rhoades, J. D., Corwin, D. L., and S. M. Lesch. 1997. Geostatistical analysis of a soil salinity data set. Advances in Agronomy, 58:241–292.

Deutsch, C. V. and Journel, A. G. 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, 2nd edition.

Dobermann, A. and J. L. Ping. 2004.Geostatistical Integration of Yield Monitor Data and Remote Sensing Improves Yield Maps. Agron. J. 96:285-297.

Drummond, S. 2006. Yield Editor 1.02 Beta Version User's Manual January, 2006. Available: http://www.ars.usda.gov/Services/Services.htm?modecode=36-22-15-00 (Accessed March 1, 2006).

Faechner, T. 2002. Probabilistic Modeling of Weed Distributions for Optimal Treatment. PhD Thesis. University of Alberta.

Faechner, T., Norrena, K., A. G. Thomas, A.G., and C. V. Deutsch 2002 A Risk-Qualified Approach to Calculate Locally Varying Herbicide Application Rates. Journal of Weed Research 42:476-485.

Fisher, R. A. 1931. Principles of Plot Experimentation in Relation to the Statistical Interpretation of the Results. Report of a Conference on the Technique of Field Experiments held at Rothamsted, May 7, 1931, pp 11-13.

Girma, K., Martin, K. L., Anderson, R. H., Arnall, D. B., K.D. Brixey, K. D., Casillas, M. A., Chung, B., Dobey, B. C., Kamenidou, S., Kariuki, S. K., Katsalirou, E., Morris, J. C., Moss, J. Q., Rohla, C. T., Sudbury, B. J., Tubana, B. S., and W.R. Raun. 2006. Mid-Season Prediction of Wheat Grain Yield Potential Using Plant, Soil, and Sensor Measurements. J. Plant Nutrition (in press.).

Griffin, T. W., Lambert, D. W., and J. Lowenberg-DeBoer. 2004. Testing for Appropriate On-Farm Trail Designs and Statistical Methods for Precision Farming. Proceedings 7th International Conference on Precision Agriculture and Other Resources Management, ASA/SSSA/CSSA, Madison. WI.

Griffin, T. W., Brown, J. P., and J. Lowenberg-DeBoer. 2005 Yield Monitor Data Analysis: Data Acquisition, Management, and Analysis Protocol. Dept. Agricultural Economics, Purdue University, Indiana.

Griffin, T. W., Florax, R. J. G. M., and J. Lowenberg-DeBoer. 2006. Field-Scale Experimental Designs and Spatial Econometric Methods for Precision Farming: Strip-Trial Designs for Rice Production Decision Making. Presented at the Southern Agricultural Economics Association Annual Meeting, Orlando, Florida.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Boil. Fertility Soils 27:315-335.

Isaaks, E. H. and R. M. Srivastava. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Journel, A. G. and C. J. Huijbregts. 1978. Mining Geostatistics. Academic Press New York.

Kaspar, T. C., Pulido, D. J., Fenton, T. E., Colvin, T. S., Karlen, D. L., Jaynes, D. B., and D. W. Meek. 2004. Relationship of Corn and Soybean Yield to Soil and Terrain Properties. Agron. J. 96:700-709.

Kravchenko, A. N., Robertson, G. P. Thelen, K. D., and R.R. Harwood. 2005. Management, Topographical, and Weather Effects on Spatial Variability of Crop Grain Yields. Agron. J. 97:514-523.

Kravchenko, A. N. and D. G. Bullock. 2002. Spatial Variability of Soybean Quality Data as a Function of Field Topography: I. Spatial Data Analysis. Crop Sci. 42:804-815.

Lauzon, J. D., Fallow, D. J., O'Halloran, I. P., Gregory, S. D. L., and A. P. von Bertoldi. 2005. Assessing the Temporal Stability of Spatial Patterns in Crop Yields using Combine Yield Monitor Data. Can. J. Plant Sci. 85:439-451.

Machado, S., Bynum, E. D., Jr., Archer, T. L., Lascano, R. J., L. Wilson, L. T., Bordovsky, J., Segarra, E., Bronson, K., Nesmith, D. M., and W. Xu 2002. Spatial and Temporal Variability of Corn Growth and Grain Yield: Implications for Site-Specific Farming. Crop Sci 42: 1564-1576.
O'Halloran, I. P., von Bertoldi, A. P., and Peterson, S. 2004 Spatial Variability of Barley (*Hordeum vulgare*) and Corn (*Zea mays* L.) yields, yield response to fertilizer N and Soil N Test Levels. Can. J. Plant Sci. 84:307-316.

Pruss, J. A., Beegle, D. B. Turgeon, A. J., Day, R. L., and R.D. Weaver. 2005. Using On-Farm Data to Validate Crop Management Recommendations and Implementation. Agron. J. 97:113-117.

Rossi, R.E. Mulla, D. H., Journel, A. G., and E. H. Franz. 1992. Geostatistical Tools for modeling and interpreting ecological spatial dependence. Ecological Monographs 62 277-314.

Simbahan, G. C., A. Dobermann, A., and J. L. Ping. 2004 Screening Yield Monitor Data Improves Grain Yield Maps. Agron. J. 96:1091-1102.

Upchurch, D. R. and W. J. Edmonds. 1991. Statistical procedures for specific objectives. In Mausbach, M. J. and Wilding, L. P., editors, Spatial Variabilities of Soils and Landforms, pages 49–71. Soil Science Society of America, Madison, Wisconsin.

Wilding L.P., Bouma, J., and D. W. Boss. 1994. Impact of Spatial Variability on Interpretive Modeling. In: Bryant R.B. and R.W. Arnold - Quantitative Modeling of Soil Forming Processes. SSSA Special Publ., No. 39: 61-75.

9.0 Appendices

Appendix 1 Descriptive statistics for selected fields with crop yield data in Alberta.

Appendix 2 Interview Protocol Utilized on Farm Operators with Grain Mapping Technology

Appendix 3 Glossary of Terms

Appendix 4 Detailed Statistical Analysis

Appendix 1 Descriptive statistics for selected fields with crop yield data in Alberta.

Field	Year	Industrial Site	Number of Yield Locations	Mean	Variance	CV	Skewness	Kurtosis
1	1997	Well	484	61.3	40.9	10.4	0.23	0.23
1	1997	Off	477	73.6	28.9	7.3	1.12	10.1
1	1997	Random	484	75.9	20.3	5.9	-1.17	6.42
1	1998	Well	484	48.2	45.6	14.0	0.62	0.12
1	1998	Off	477	44.4	36.4	13.6	0.33	3.9
1	1998	Random	484	38.4	42.2	16.9	-0.13	1.6
1	2000	Well	484	53.2	48.2	13.0	-0.62	0.22
1	2000	Off	477	53.7	59.3	14.3	-0.18	-0.12
1	2000	Random	484	50.2	33.5	11.5	0.47	-0.25
1	2003	Well	484	39.8	12.8	9.0	0.18	0.25
1	2003	Off	477	40.5	19.6	10.9	-0.18	-0.56
1	2003	Random	484	33.9	28.1	15.6	0.6	0.45
2	2005	Well	441	52.4	35.2	11.3	-0.96	0.94
2	2005	Off	458	52.7	41.5	12.2	-1.02	1.05
2	2005	Random	441	46.2	36.7	13.1	-1.21	1.89
3	2000	Well	441	38.0	233.7	40.3	-0.23	-0.86
3	2000	Off	450	43.3	115.8	24.9	-0.79	0.61
3	2000	Random	441	47.3	49.9	14.9	-0.64	-0.29
3	2004	Well	441	59.0	51.1	12.1	-0.21	1.71
3	2004	Off	450	57.8	91.4	16.5	-0.56	0.61
3	2004	Random	441	60.8	278.2	27.5	-0.53	-0.93

Appendix 1 Descriptive statistics for selected fields with crop yield data in Alberta.

Field	Year	Industrial Site	Number of Yield Locations	Mean	Variance	CV	Skewness	Kurtosis
4	2002	Pipeline	340	16.5	7.2	16.3	0.3	-0.90
4	2002	Off	340	16.5	8.6	17.7	0.52	-0.74
4	2002	Random	340	17.0	6.1	14.5	0.55	-0.07
5	2003	Borrow Pit	486	65.6	99.4	15.2	1.7	5.8
5	2003	Off	476	91.4	371.9	21.1	0.09	-0.20
5	2003	Random	486	111.6	262.6	14.5	-0.42	1.3
5	2005	Borrow Pit	486	71.2	80.9	12.6	-0.48	2.7
5	2005	Off	486	75.1	86.5	12.4	-0.33	2.4
5	2005	Random	486	72.4	76.9	12.1	-0.42	2.6
6	2005	Pipeline	382	60.2	135.7	19.3	-0.33	0.07
6	2005	Off	382	59.5	105.1	17.2	-0.13	0.65
6	2005	Random	382	55.5	101.6	18.2	-0.41	0.39
7	2004	Pipeline	335	47.1	20.8	9.7	-0.28	0.94
7	2004	Off	335	45.7	21.3	10.1	-0.27	1.31
7	2004	Random	335	45.3	24.4	10.9	-0.12	1.33
8	2004	Pipeline	450	50.9	27.1	10.2	-0.19	2.31
8	2004	Off	450	47.9	33.2	12.0	-0.45	3.59
8	2004	Random	450	44.6	30.7	12.4	-0.97	3.19

Appendix 2 Interview Protocol Utilized on Farm Operators with Grain Mapping Technology.

Interview Protocol

Upon identifying potential prospects, contact by telephone, identify yourself, and indicate that this is not a survey, but that you would like to ask a few questions for a potential applied research project.

- Do you have a combine yield monitor with GPS?
- Do you have any reclaimed industrial lands, such as: upstream oil and gas wellsites, road borrow sites, pipelines, sand and gravel pits, battery sites, plant sites (e.g. compressor sites)?
    - If yes, what is the status of each site (e.g. reclaimed? certified?)?
    - What year was it reclaimed or certified?
- Would you be interested in participating in a study to verify well-site/ industrial land reclamation using combine yield monitors with GPS as a tool to collect yield data?
- Have you used your combine yield monitor with GPS turned on when harvesting the field with your reclaimed industrial site?
    - If yes, do you have the data downloaded on your office computer, and how many years' of data do you have?
- What is/ are the location(s) of any reclaimed industrial site(s) (e.g. legal land description(s), municipality)?
- What crops were grown on the site before and after it was reclaimed?
- How many acres does the disturbed site cover?
- Do you have any soil tests or other diagnostic soil or crop indicator information available from this site or field?
- What type of harvest equipment and computer software do you have (e.g. combine, yield monitor brand, GPS and office computer software for analyzing the yield data (e.g. John Deere (JD Office), Ag Leader (SMS), Case IH (AFS), New Holland (PFS), Caterpillar (Fieldstar))?
- Please describe the performance history of the site (e.g. type of use, company, issues, problems, relationships)?
- Do you have anything else to add?

Appendix 3 Glossary of Terms

Anisotropy – directional dependence of a parameter such as the variogram range.

Bootstrapping - a statistical method based on repeated random sampling with replacement from an original sample to provide a collection of new samples, from which the sampling variance can be estimated.

Borrow pit - a site for the extraction of aggregate materials (sand, gravel, or soil) over a limited period, for exclusive use in a specific construction project, which will usually be close to or contiguous with the site.

Cokriging - kriging applied with more than one data type.

Filtering – technique to transform data due to errors in measurement.

First order moment – mean.

Geometric - the range changes with direction; there is a single sill, but the variogram reaches the sill in a shorter lag distance along a certain direction.

Geo-referenced - referencing in space of the location of a point using a predefined coordinate system such as a latitude/longitude.

Geostatistics – study of natural resource phenomena that vary in space and/or time.

Grid node – rectangular area formed by two sets of regularly spaced lines and where these lines intersect represents a node.

Isotropic – showing no directional trend or anisotropy.

Kurtosis - a measure of peakedness around the mode.

Lag distance - is a distance within which any two samples (a pair) is taken for variogram calculation.

Monte Carlo simulation - a technique of drawing or sampling from a probability distribution using a random number sequence with multiple possible values.

Nested – refer to multiple variogram structures that are combined to model the spatial variation in complex settings.

Normalized difference vegetation index (NDVI) - a remotely sensed satellite-derived measure of the radiation absorbed by chlorophyll during plant photosynthesis that is a correlate of soil moisture, vegetation biomass and productivity.

Nugget – errors in microscale measurement or sampling errors that are smaller than the sampling interval or both.

Ordinary kriging – spatial linear regression when the mean is constant but unknown.

Random function – statistical model of a variable distributed in space with a probability distribution.

Range - the separation distance over which spatial dependence is apparent and the distance at which the variogram reaches the sill.

Second order moment – variance.

Sequential Gaussian simulation – building alternative, equally probable models of the spatial distribution of a variable like crop yield that has a normal distribution.

Sill – total variance.

Simple kriging - spatial linear regression when the constant mean is known in the area.

Skewness – departure of a probability density function from being symmetric.

Spatial variability - differences in observed attribute(s) that are noted between locations in a field.

Stationarity – a decision that the statistics do not depend on location over an area and the variogram model is assumed to be applicable over the entire area.

Structure – a variable that has spatial correlation.

Variogram – tool to measure the variance between data as a function of distance for describing spatial correlation.

Variography – iterative procedure of calculating, interpreting and modeling variograms.

Zonal anisotropy – directional variability of a variable. The variogram for a variable with zonal anisotropy has different sills for different directions.

Appendix 4 Detailed Statistical Analysis

1.0 Introduction

1.1 Crop Yield Data Filtering

Preparing crop yield data for spatial analysis is often referred to as "data cleaning" or "data filtering". It involves removing crop yield observations that are known to be erroneous due to combine machine dynamics and operator behavior. These data filtering procedures are not a modification or manipulation of the data (Simbahan et al. 2004; Griffin et al. 2005).

With yield monitor data, many observations have erroneous yield values due to harvester velocity, velocity change, and maximum yield. With harvester yield data, errors also arise from start and stop delays for beginning and ending of passes. The ramping up and ramping down effects of the harvester yield monitor has adverse effects on yield measurements. Flow delay caused by inaccurate assignment of yield measurement to GPS coordinates can occur. Simbahan et al. 2004 provides a description of the procedures that can filter crop yield data. Their procedures removed 13 to 20% of the soybean and corn yield values. Griffin et al. 2005 provides a manual for implementing these filtering techniques on crop yield data.

1.2 Crop Yield Spatial Variability

Spatial variation has been studied using geostatistics, which is a set of tools where the assumptions of sample independence and homogeneity are removed (Upchurch et al., 1991). These tools measure the degree of dependence of samples. They have been applied extensively in mining and petroleum exploration with the objective of quantifying resources, but on a limited basis in soil science (Bourgault et al., 1997; Deutsch and Journel, 1998; Goovaerts, 1997; Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978).

A geostatistical analysis of crop yield could consist of: 1) exploratory data analysis using descriptive statistics, 2) spatial continuity of crop yield, 3) map building based on simulation, and 4) decision making in the presence of spatial uncertainty (Deutsch and Journel, 1998; Goovaerts, 1997; Isaaks and Srivastava, 1989). Spatial continuity of variables has led to the theory of regionalized variables. A random function (RF) is a set of random variables, (RV) $Z(\mathbf{u})$ defined over multiple locations, $\mathbf{u}$. The mathematical representation of this spatial variability may be provided by a random function concept (Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978).

Spatial modeling begins with determining the variogram parameters for a particular model. Variogram analysis can be used to compare crop yields from different locations in a field. Crop yield at the small-scale (variogram values at short lag distances and variogram slopes near the origin) were determined to be more sensitive to the small-scale variations in growth conditions due to soil and topographical differences using variograms (Kravchenko et al., 2005). Dobermann and Ping, 2004 created interpolated crop yield maps from harvest monitor yield data and remote sensed vegetation indices. Geostatistical procedures including the calculation of variograms were implemented to provide parameters for ordinary kriging, cokriging and simple kriging. Kriging is an interpolation technique that allows the estimation of crop yield values at unsampled locations in the field.

Variogram models can be fit for mapping. Kravchenko and Bullock, 2002 used spherical and exponential variogram models for soybean oil and protein concentration in five fields in central Illinois. Kaspar et al. 2004 determined that the spherical variogram was the best fit for the empirical terrain and soil data in the central Illinois study. The relationship between soybean and corn yields to soil variables such as pH, organic C, N, K, P and Zn, electrical conductivity, elevation, and slope was compared over 11 years in this investigation. O'Halloran et al., 2004 modeled barley and corn yields with exponential variograms for 1996 and 1997 in Ontario.

Simulation or sequential Gaussian simulation has been applied to electromagnetic data from soils (Bourgault et al., 1997) and to quantify uncertainty in weed density data (Faechner, 2002; Faechner et al. 2002). Griffin et al., 2004 used Monte Carlo simulation to evaluate the accuracy of two regression methods based on the spatial correlation of crop yield data. Simulation generates models, and differences between models can be used to provide a confidence interval.

1.3 Variography

When crop yield data are correlated in space, they will have a lower variance or γ at smaller separation distances than at larger separation distances (Figure A1). The nugget represents unexplained variation due to measurement error and/or sources of spatial variation at very small distances. As separation distance increases beyond some distance called the range, the variance values often stay constant at a value known as the sill (Deutsch and Journel, 1998).

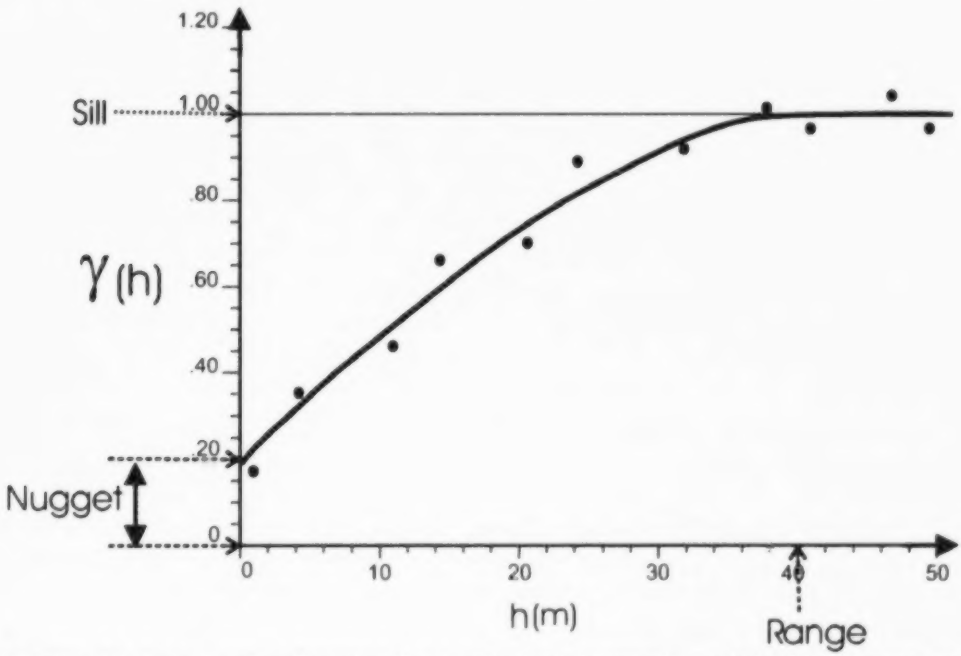


Figure A1 A variogram function for a spherical, isotropic model with a nugget variance (0.20). The dots represent calculated points of the experimental variogram while the solid line is the model variogram.

A variogram model can be constructed as a sum of variogram functions:

$$\gamma(h) = \sum_{i=1}^{nst} C_i \Gamma_i(h)$$

where (h) is a variogram model for direction and distance vector h, *nst* is the number of variogram functions or "nested structures", C_i. $i = 1,\ldots nst$ are the variance contributions for each nested structure, and $\Gamma_i, i = 1 \ldots nst$ are the variogram functions. The variance contributions of each nested structure sum to one that is the sill. Each nested structure could have different range parameters in different directions.

Variogram models are often fit to experimentally calculated points. Each variogram model has a lag distance *h* as a parameter. This distance is calculated by breaking down the distance vector **h**, into its 3 principal directions, vertical (h_{vert}), major horizontal ($h_{h\text{-}major}$) and minor horizontal ($h_{h\text{-}minor}$):

$$h = \sqrt{\left(\frac{h_{vert}}{a_{vert}}\right)^2 + \left(\frac{h_{h-major}}{a_{h-major}}\right)^2 + \left(\frac{h_{h-minor}}{a_{h-minor}}\right)^2}$$

The distance range parameters, *a*, in a_{vert}, $a_{h\text{-}major}$ and $a_{h\text{-}minor}$ correspond to each direction for the nested variogram structure. Geometric anisotropy results from variograms that have the same sill but different ranges whereas zonal anisotropy occurs when the sill changes with direction. Each structure may have its own anisotropy (Deutsch and Journel, 1998).

Typical variogram models include linear, spherical, exponential, Gaussian, and Power law. A model for an empirical variogram is the spherical and Gaussian model given by:

$$\gamma(h) = C * Sph(h) = \begin{cases} C*[1.5(h) - 0.5(h)^3] & \text{if } h \le 1 \\ C & \text{if } h \ge 1 \end{cases}$$

$$\gamma(h) = \begin{cases} C * [1 - \exp(-3h)^2] & \text{if } h \le 1 \\ C & \text{if } h \ge 1 \end{cases}$$

where *C* is the variance contribution for each nested structure. The range is 1 if *h* is normalized as above. Directional trends or anisotropy in crop yield data are common (Deutsch and Journel, 1998).

1.4 Spatial Averages

Geostatistical simulation is an approach to spatial modeling that attempts to reproduce the range of crop yield values present in the original data, as well as the spatial variability described by the variograms. Instead of producing a single, average estimate, a geostatistical simulation produces several alternative, equiprobable models. These models represent spatial averages for crop yield data. Simulation can be used to produce many models, for which a statistical distribution of crop yield estimates with confidence intervals can be determined. The confidence intervals establish the probability of crop yield estimates falling within a certain range (i.e. a 95% probability that all crop yields fall between 50 and 75 bushels per acre).

This approach was applied to estimate crop yields with confidence intervals using the nonparametric procedure bootstrapping (Pruss et al. 2005). Bootstrapping is a resampling method to create new crop yield values for making inferences. Similarly, Griffin et al. 2004 used Monte Carlo simulation to evaluate different levels of spatial autocorrelation for crop yields. Monte Carlo simulation is stochastic, meaning it is based on the use of random numbers and probability statistics to investigate problems.

Sequential Gaussian simulation (Deutsch and Journel, 1998; Faechner, 2002; Goovaerts, 1997) is used to generate unconditional, simulated crop yield values using the mean and variogram of the crop yield data collected from the field. In unconditional simulation, we simulate the field with no reference to the actual sample, i.e. the data we have. This is mainly to visualize a random field as modeled by a variogram, not for prediction. Unconditional simulation reproduces global statistics, such as the variogram, and preserves the mean and covariance structure.

2.0 Methods

2.1 Spatial Analysis

On completion of filtering of crop yield data (Figure A2), the data is analyzed using geostatistical procedures to produce variograms for the industrial and off-site areas, using the public domain software GSLIB (Deutsch and Journel, 1998).

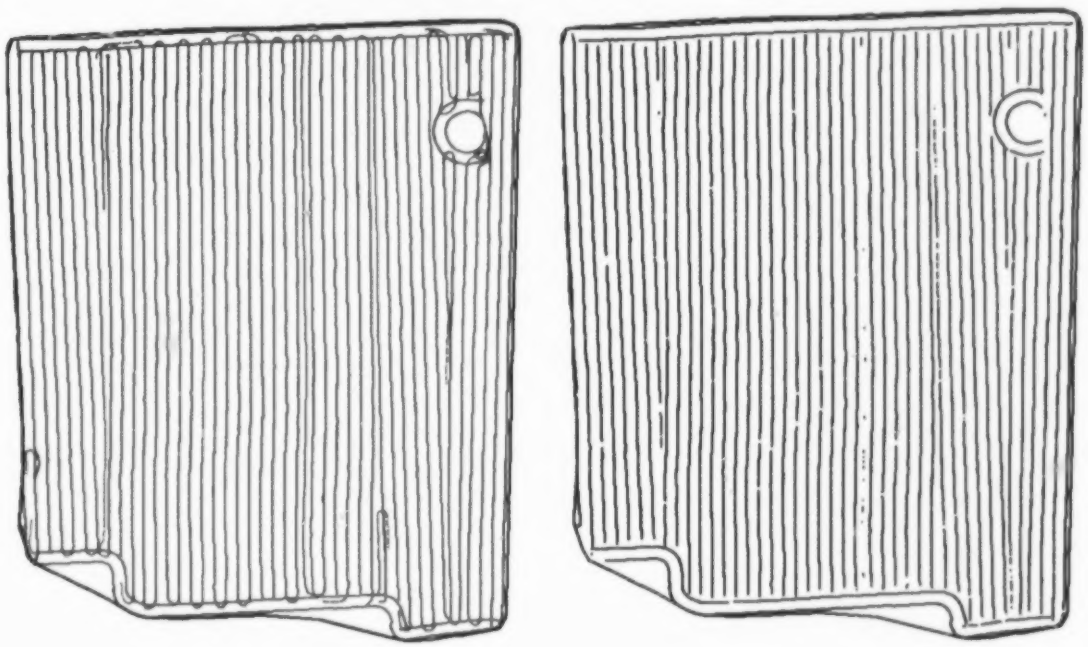


Figure A2 Filtering of combine barley yield data for field 9 in 2003. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the ends of the field where the combine turned to begin harvesting the crop for erroneous yield data.

A mathematical representation of spatially distributed variables is the "random function" concept (Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978). Uncertainty in the crop yield value, $z(\mathbf{u})$, at location $\mathbf{u}$ is modeled with a random variable (RV) $Z(\mathbf{u})$. The random variable, $Z(\mathbf{u})$, is location dependent. A set of RVs is called a random function (RF), denoted by $\{Z(\mathbf{u}), \mathbf{u} \in \text{field area A}\}$. The variogram for two locations $\mathbf{u}_1$ and $\mathbf{u}_2$ is defined as:

$$\gamma(h) = \tfrac{1}{2} \text{ var}\{[Z(\mathbf{u}_1) - Z(\mathbf{u}_2 + h)]^2\},$$

where $\gamma(h)$ is the variance between $\mathbf{u}_1$ and $\mathbf{u}_2$, separated by a distance of $\mathbf{h}$, referred to as the lag distance.

The mean $m(\mathbf{u})$ is a measure of central tendency and the variance $\sigma^2(\mathbf{u})$ is a measure of spread. The covariance measures correlation or similarity of two RVs whereas the variogram measures variability or dissimilarity of two RVs.

The first and second order moments depend on location $\mathbf{u}$. Statistical inference may require many repetitive models to calculate distributions and expected values. Such replication is unavailable and we must pool data from different locations. Then the $F\{Z(\mathbf{u}), \mathbf{u} \in \text{field area A}\}$ is said to display first order stationarity since its univariate cdf and first order moment, the mean, are invariant. A decision of stationarity allows inference (Journel and Huijbregts, 1978) and provides for the assumption of "spatial homogeneity" over a study area. The decision of stationarity for developing reliable inferences lies with the practitioner. If more data becomes available or the scale of the study changes, the decision of stationarity may change.

When crop yield values are correlated in space, they will have a lower variance at smaller separation distances than at larger separation distances (Figure A1). The nugget represents unexplained variation while the range is the separation distance; the variance values often stay constant at a value known as the sill (Deutsch and Journel, 1998).

Variogram models are fit to experimentally calculated crop yield values. Each variogram model has a lag distance, $\mathbf{h}$, as a parameter. This distance is calculated by breaking down the distance vector $\mathbf{h}$ into its vertical, major horizontal and minor horizontal directions. Geometric anisotropy results from variograms that have the same sill but different ranges, whereas zonal anisotropy occurs when the sill changes with direction. Typical variogram models include linear, spherical, exponential, Gaussian, and Power law.

Model variograms are assumed to represent experimental variograms for crop yield values. Curve fitting of a model variogram relies on expert judgment that is subject to interpretation. Geographic location information that indicates where soil nutrient data is sampled is subject to error during collection.

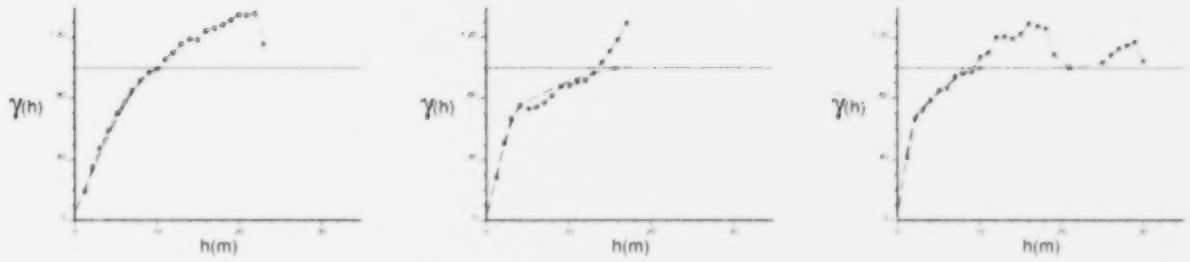


Figure A3 Field 1 experimental and model variograms for the well site area, off-site around the well site and random off-site area (left to right). The red dots are the experimental variogram, while the black dashes are the model variogram and the **h** or lag distance is expressed as meters.

Variogram models are calculated from the crop yield data for three different areas in the field. The experimental and model variograms for the 1997 barley yields from field 1 are illustrated in Figure A3. The areas include the industrial site, which may be a well site, pipeline, or borrow pit; an area immediately surrounding or beside the industrial site, as is the case for the well site, borrow pit or pipeline; and an area whose location is randomly determined in the field. These areas are illustrated in Figure A4. Barley crop yield values for 1997 from field 1 for the three areas are provided in Figure A5.

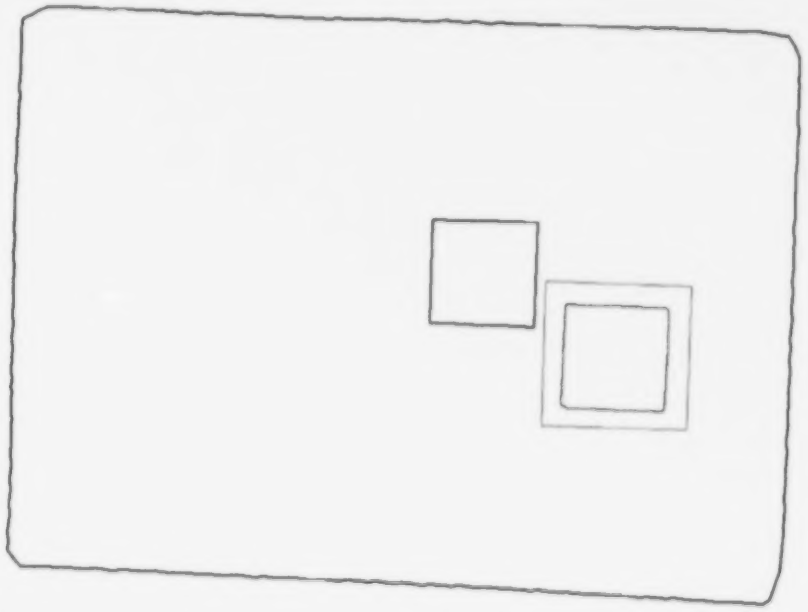


Figure A4 Field 1 illustrates the 3 different areas for calculating variogram models. The red rectangle defines the well site area; the green rectangle is the off-site around the well site, the blue rectangle is the random off-site area and the black rectangle is the field boundary.

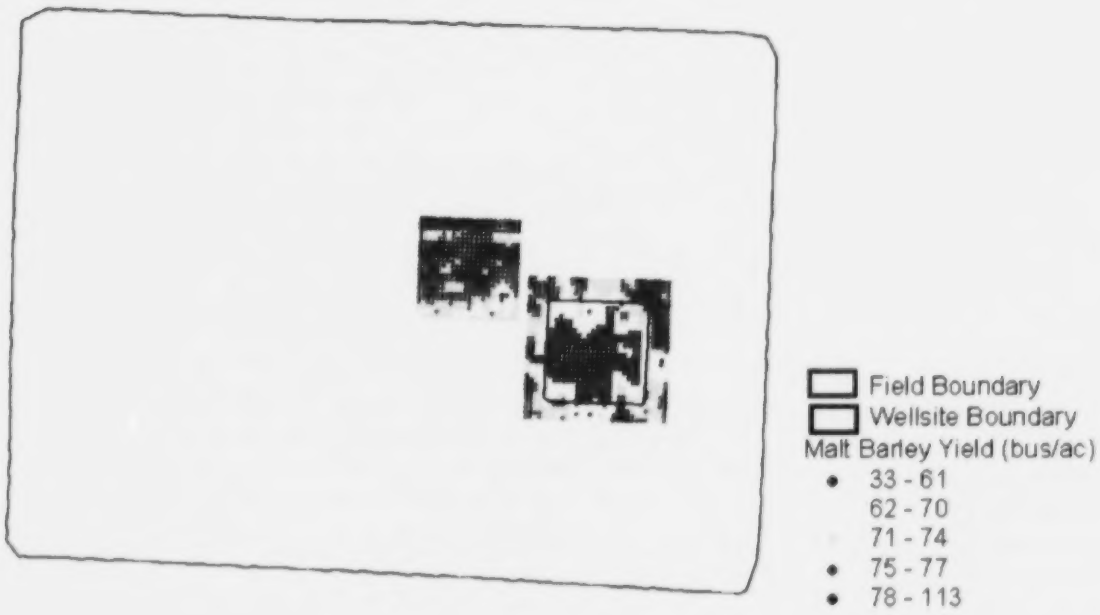


Figure A5 Field 1 illustrates the 3 different areas with barley yield values for each area in 1997. The black rectangle distinguishes the inside area, or well site, from the off-site area that surrounds the well site.

2.2 Spatial Averages

The basic steps in implementing sequential Gaussian simulation include: 1) generating a random path through the grid nodes; 2) visiting the first node along this path and kriging to estimate a mean and standard deviation for the variable at that node based on surrounding data values; 3) selecting a value at random from the corresponding normal distribution and setting the variable value at that node to that number; and, 4) visiting each successive node in the random path and repeating the process using previously simulated nodes as data values (Deutsch and Journel, 1998).

Sequential Gaussian simulation is used to generate 101 unconditional, simulated models using the mean, variogram and crop yield data from an area in a field. For each of the areas described in Table A1, there are models created for the industrial site, off-site around the industrial site and randomly located off-site in the field. Three models for field 1 for each area are displayed in Figure A6. These models are used to generate a spatial average that is a mathematical mean computed using data from grid points of the model. Confidence intervals at the 95% level of significance are calculated from each model (Figure A7). A cumulative probability plot of the confidence intervals for each model graphically illustrates all the data values. Therefore, if the confidence intervals from an area do not overlap with those of another, the crop yields of the two areas are significantly different (Figure A7).

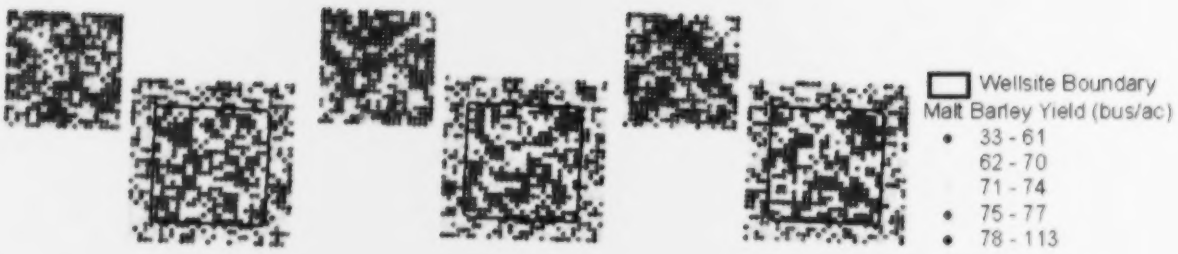


Figure A6 Field 1 illustrates 3 of 101 models of barley yield values for the 3 different areas in 1997. The black rectangle distinguishes the inside area, or well site, from the off-site area that surrounds the well site.

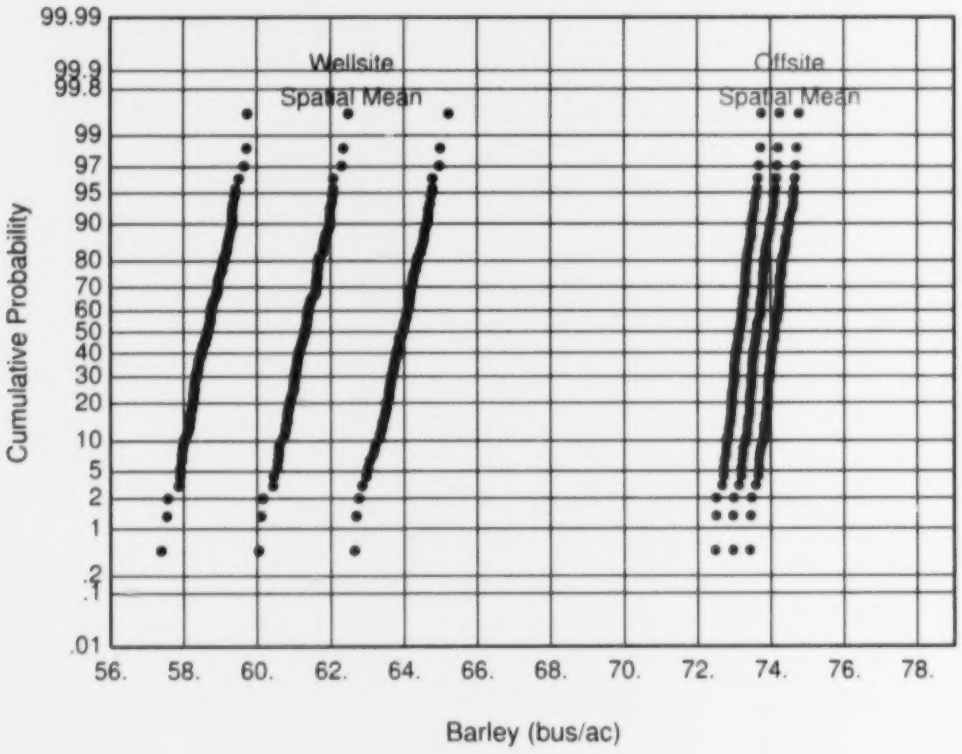


Figure A7 Confidence intervals for barley yield (bushels/acre) from field 1 in 1997 are significant at the 95% level. On either side of a spatial mean is each mean's confidence interval. There are 101 spatial means of crop yield from the well site and off-site areas. The comparison in confidence intervals between the well site (black dots) and the off-site area surrounding the well site (red dots) indicates that the off-site crop yield is significantly higher compared to the well site crop yield.

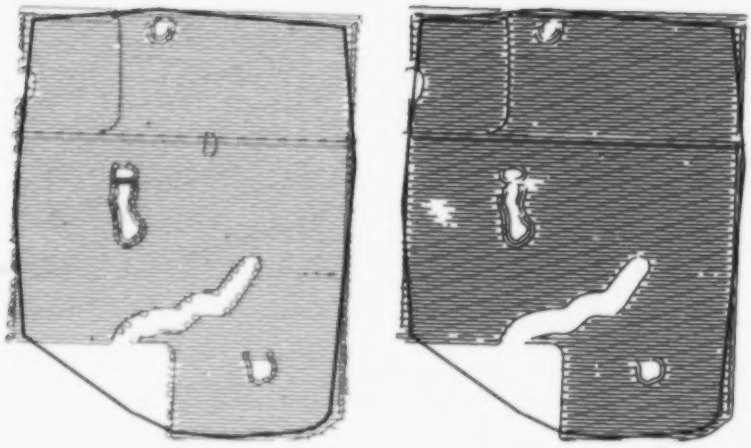


Figure A8 Filtering of combine pea yield data for field 2. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the ends of the field for erroneous yield data in the unfiltered yield map where the combine turns to begin harvesting the crop.

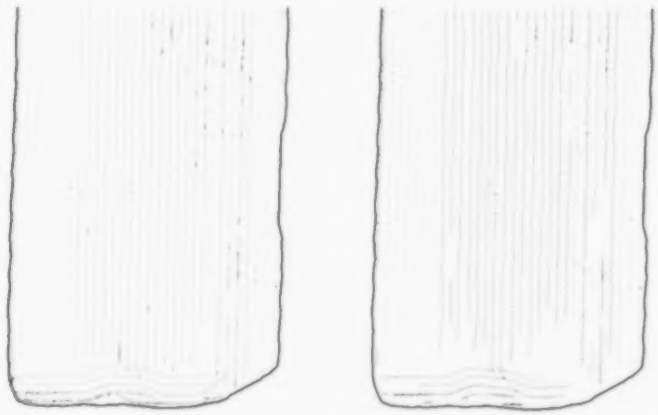


Figure A9 Filtering of combine pea yield data for field 3 in 2000. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the ends of the field for erroneous yield data in the unfiltered yield map where the combine turns to begin harvesting the crop.

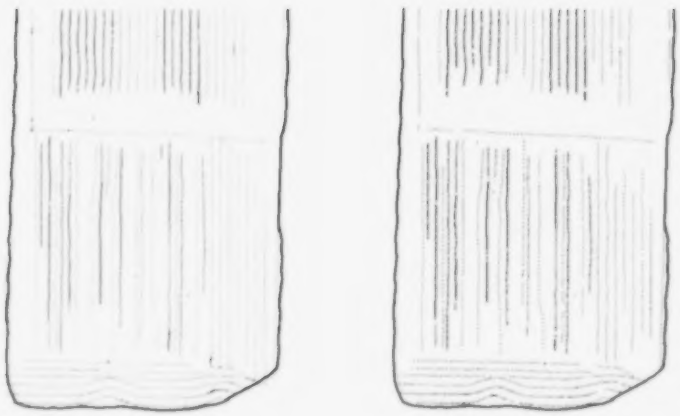


Figure A10 Filtering of combine pea yield data for field 3 in 2004. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the unfiltered yield map outside the field boundary.

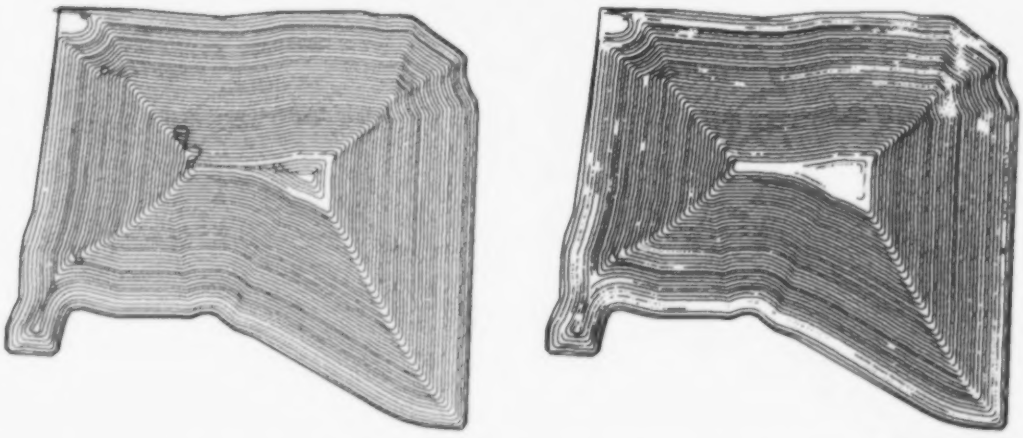


Figure A11 Filtering of combine wheat yield data for field 4. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the center of the unfiltered yield map.

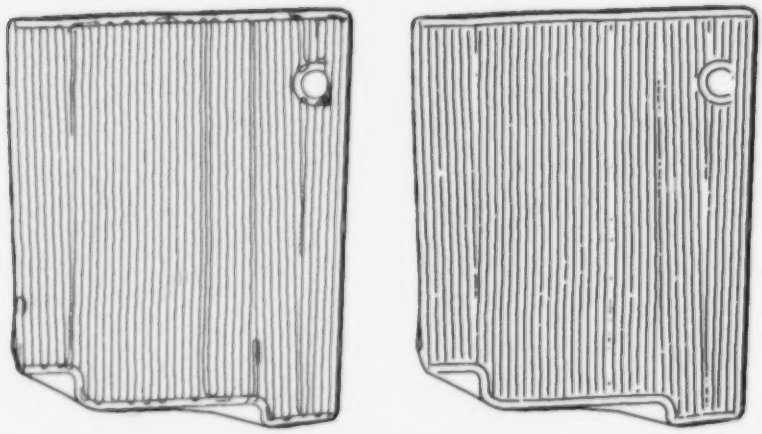


Figure A12 Filtering of combine barley yield data for field 5 in 2003. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the upper, right corner of the unfiltered yield map.

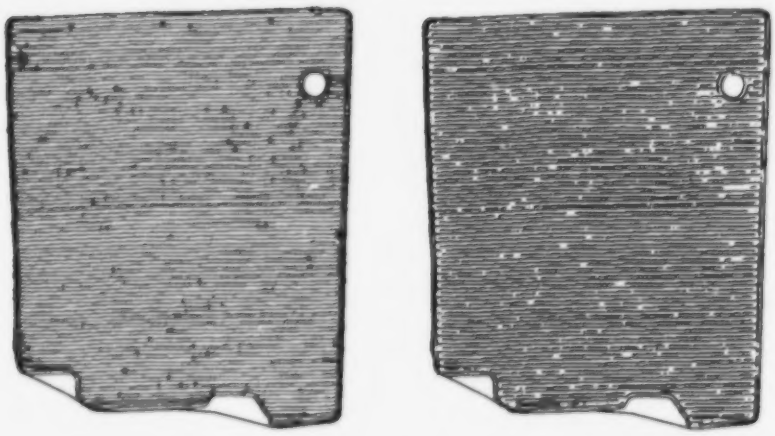


Figure A13 Filtering of combine canola yield data for field 5 in 2005. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data (red dots) throughout the unfiltered yield map.

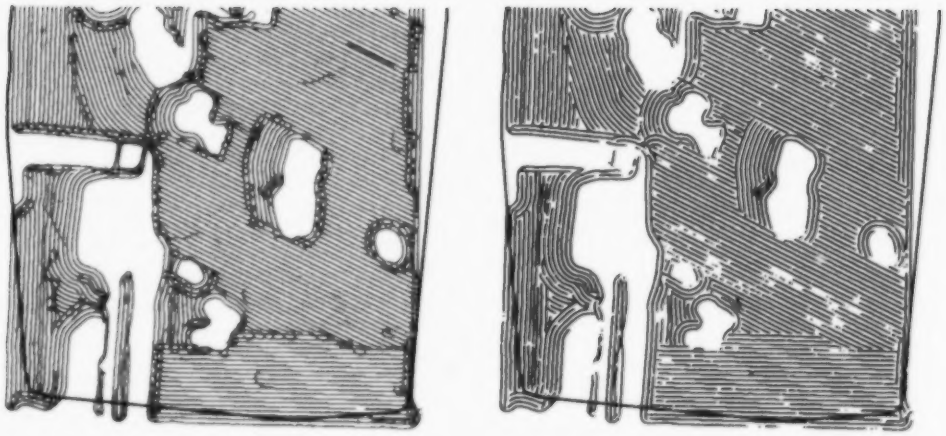


Figure A14 Filtering of combine wheat yield data for field 6. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the headlands and around the depressional areas of the unfiltered yield map. The depressional areas appear as white holes in the map.

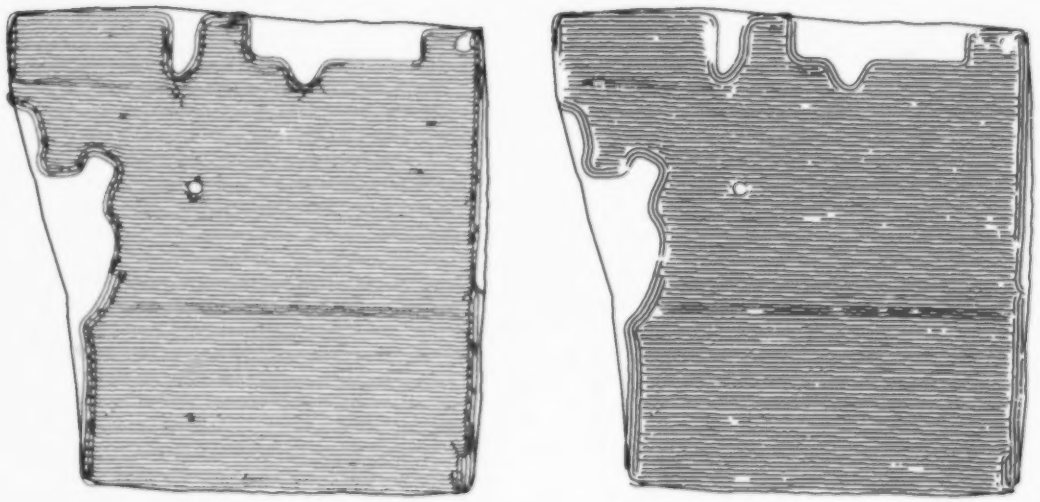


Figure A15 Filtering of combine canola yield data for field 7. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the headlands of the unfiltered yield map.

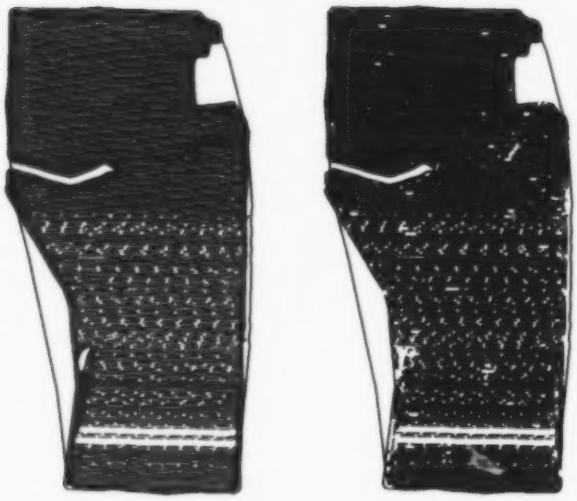


Figure A16 Filtering of combine canola yield data for field 8. The unfiltered yield data is represented on the left while the filtered yield data is on the right. The black outline represents the field boundary. Note the erroneous yield data in the headlands of the unfiltered yield map.

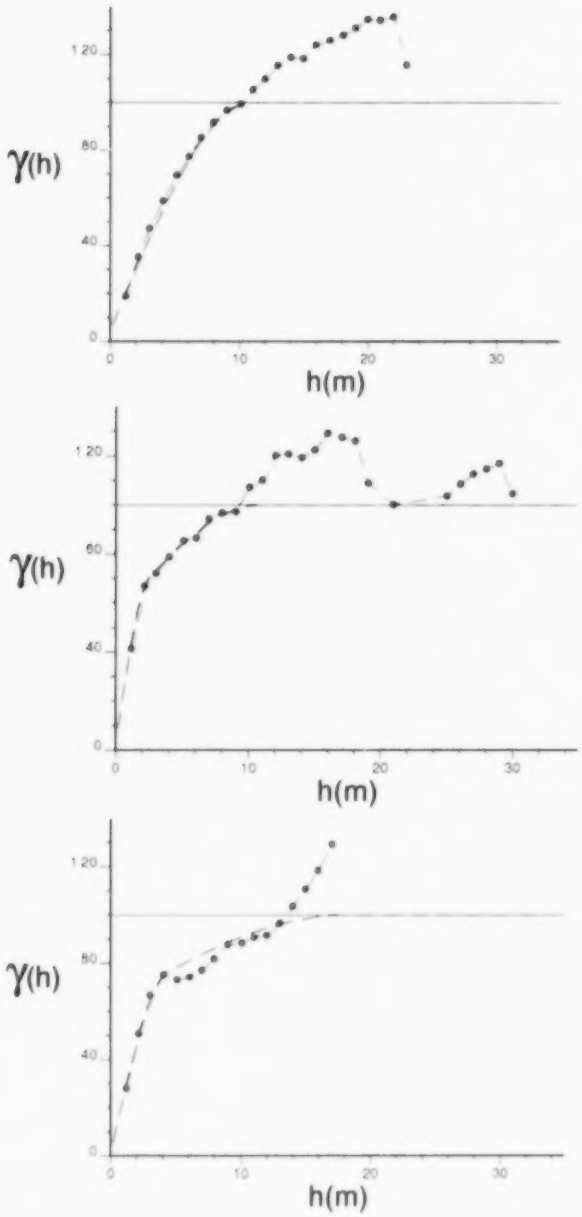


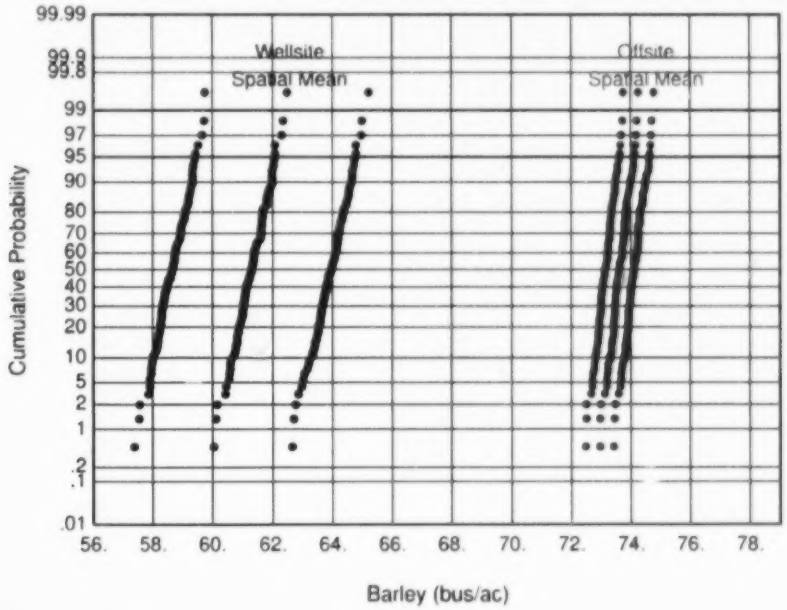
Figure A17 Field 1 experimental and model variograms for the well site area, off-site around the well site and random off-site area (top to bottom). The red dots are the experimental variogram while the black dashes are the model variogram.

Table A1 Spatial parameters for isotropic variograms for yield observations sampled at eight fields in central and northern Alberta. Continuity refers to the variance contribution of each nested structure. The variograms are modeled with nested spherical or exponential structures.

Industrial Site	Model	Nugget	Continuity 1	Range 1 (m)	Continuity 2	Range 2 (m)
Field 1 - 1997						
Well site	Spherical	0.05	0.95	11.0	---	---
Off-site	Spherical	0.03	0.52	2.5	0.45	11.0
Random Off-site	Spherical	0.02	0.61	4.5	0.37	18.0
Field 1 -1998						
Well site	Exponential	0.07	0.93	12.0	---	---
Off-site	Exponential	0.01	0.58	3.0	0.41	15.0
Random Off-site	Exponential	0.05	0.71	3.5	0.24	27.0
Field 1 - 2000						
Well site	Spherical	0.05	0.35	5.0	0.60	15.0
Off-site	Spherical	0.01	0.35	4.0	0.64	26.0
Random Off-site	Spherical	0.02	0.98	12.0	---	---
Field 1 - 2003						
Well site	Spherical	0.01	0.55	4.5	0.44	14.0
Off-site	Exponential	0.13	0.46	4.0	0.41	27.0
Random Off-site	Spherical	0.02	0.98	12.0	---	---
Field 2 - 2005						
Well site	Spherical	0.10	0.25	3.0	0.65	19.0
Off-site	Spherical	0.06	0.23	3.0	0.71	20.0
Random Off-site	Spherical	0.06	0.65	7.0	0.29	12.0
Field 3 - 2000						
Well site	Spherical	0.01	0.36	2.5	0.63	12.5
Off-site	Spherical	0.02	0.52	3.0	0.46	24.0
Random Off-site	Spherical	0.04	0.96	11.0	---	---
Field 3 - 2004						
Well site	Spherical	0.08	0.92	5.5	---	---
Off-site	Exponential	0.05	0.95	2.5	---	---
Random Off-site	Spherical	0.05	0.62	2.2	0.33	16.0
Field 4 - 2002						
Pipeline	Spherical	0.02	0.23	3.0	0.75	30.0
Off-site	Spherical	0.08	0.92	39.0	---	---
Random Off-site	Spherical	0.02	0.25	3.0	0.73	65.0
Field 5 - 2003						
Borrow Pit	Spherical	0.02	0.45	6.0	0.53	15.0
Off-site	Exponential	0.08	0.92	6.0	---	---
Random Off-site	Spherical	0.02	0.98	5.0	---	---
Field 5 - 2005						
Borrow Pit	Spherical	0.40	0.42	2.0	0.18	9.0
Off-site	Spherical	0.80	0.20	4.0	---	---
Random Off-site	Spherical	0.40	0.60	2.5	---	---
Field 6 - 2005						
Pipeline	Spherical	0.17	0.48	6.5	0.35	35.0
Off-site	Spherical	0.20	0.52	5.5	0.28	30.0
Random Off-site	Spherical	0.04	0.57	3.5	0.39	42.0

Industrial Site	Model	Nugget	Continuity 1	Range 1 (m)	Continuity 2	Range 2 (m)
Field 7 - 2004						
Pipeline	Spherical	0.35	0.40	2.5	0.10	12.0*
Off-site	Spherical	0.10	0.60	2.0	0.30	28.0
Random Off-site	Spherical	0.42	0.47	4.5	0.11	20.0
Field 8 - 2004						
Pipeline	Spherical	0.15	0.72	2.5	0.13	55.0
Off-site	Spherical	0.25	0.49	2.5	0.10	14.0**
Random Off-site	Spherical	0.30	0.37	2.5	0.19	7.0***

*Extra structure with range of 20 m and continuity of 0.15.
**Extra structure with range of 70 m and continuity of 0.16.
***Extra structure with range of 23 m and continuity of 0.14.



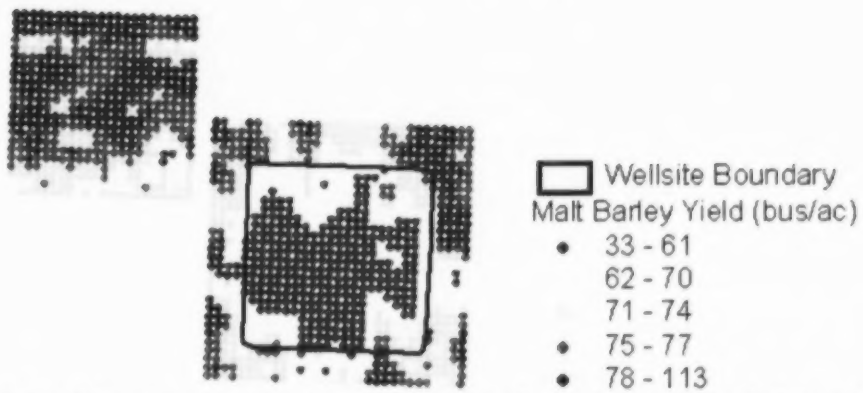


Figure A18 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 1 in 1997 significant at the 95% level for 101 models comparing the well site to the offsite area (top figure). The center set of dots is the spatial average while the dots on either side of the

spatial average represent confidence intervals. Barley yield map from field 1 for the well site, offsite and random site (bottom figure).

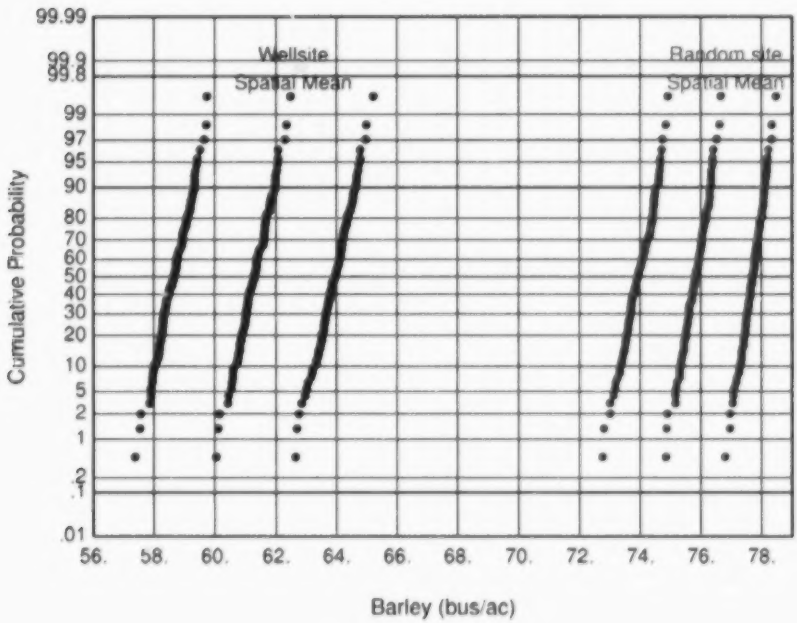
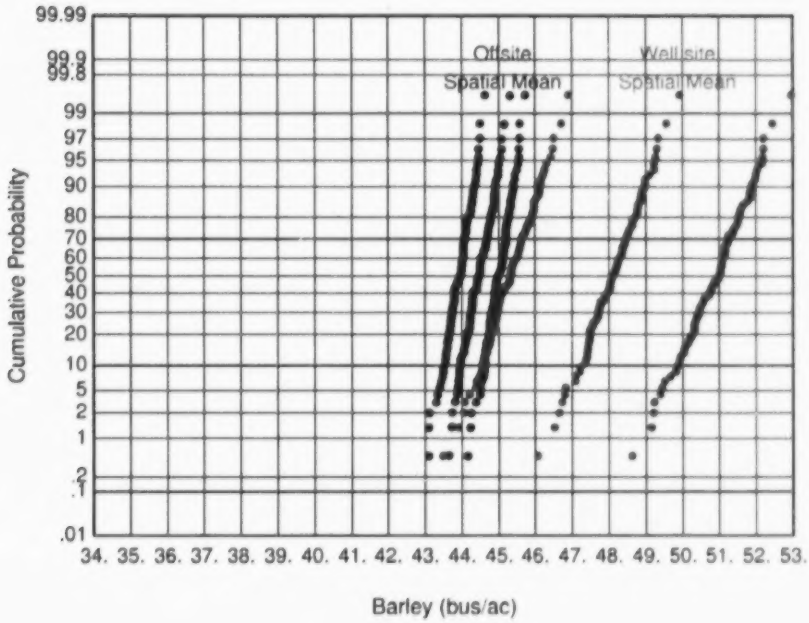


Figure A19 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 1 in 1997 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



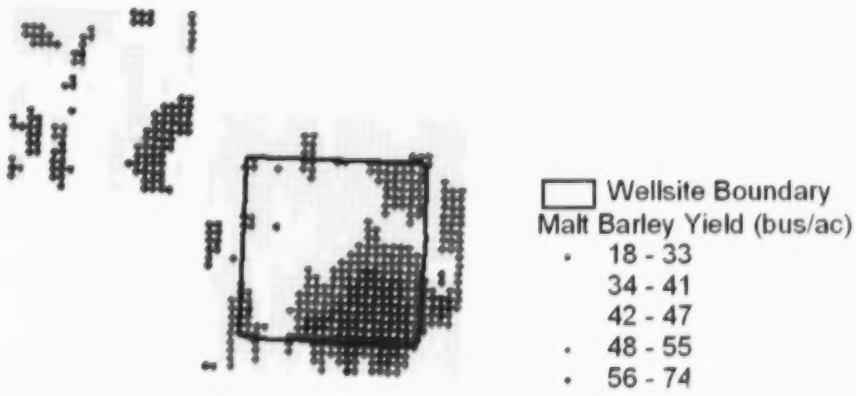


Figure A20 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 1 in 1998 significant at the 95% level for 101 models comparing the well site to the offsite area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Barley yield map from field 1 for the well site, offsite and random site (bottom figure).

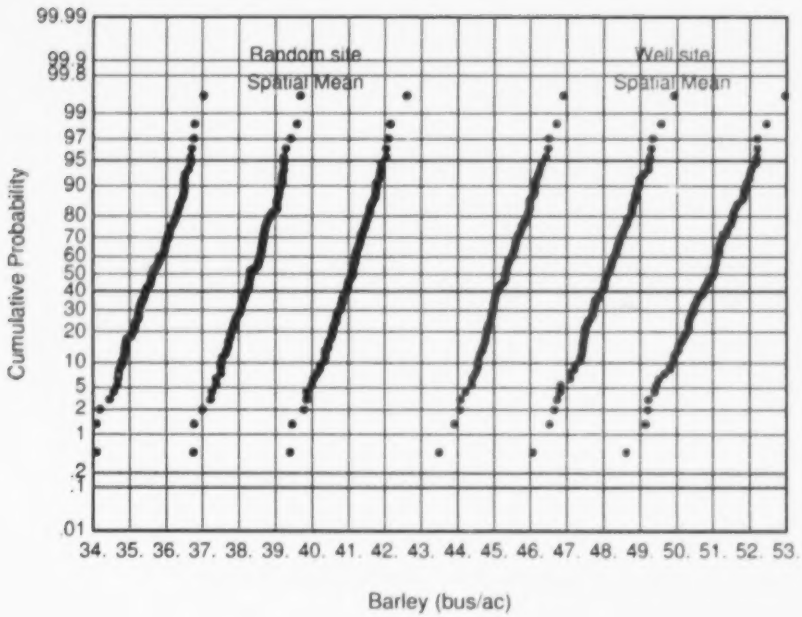
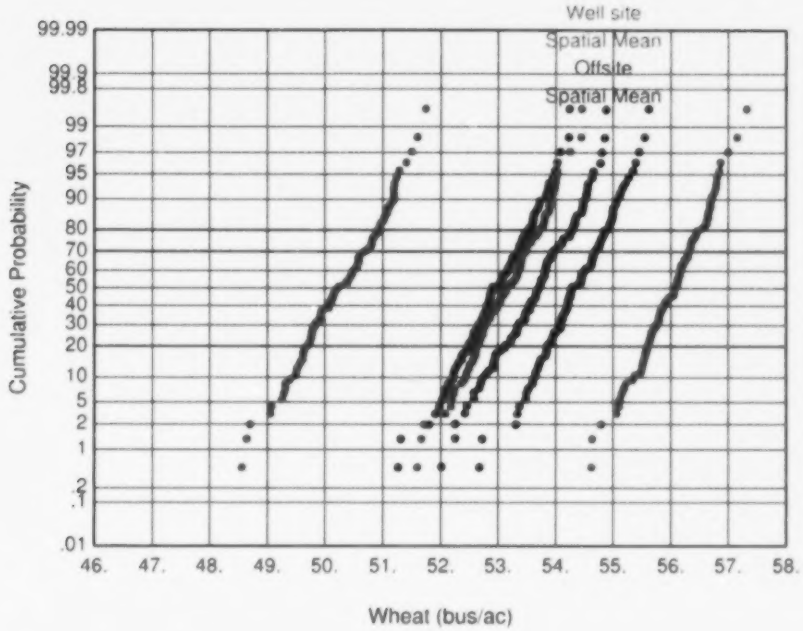


Figure A21 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 1 in 1998 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



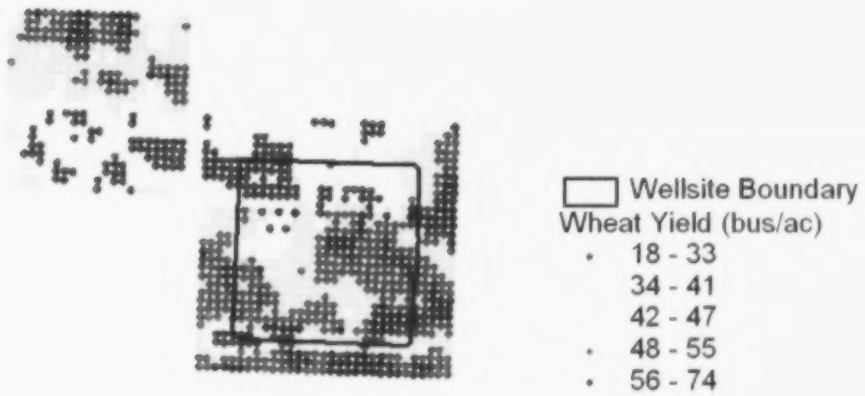


Figure A22 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 1 in 2000 significant at the 95% level for 101 models comparing the well site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Wheat yield map from field 1 for the well site, offsite and random site (bottom figure).

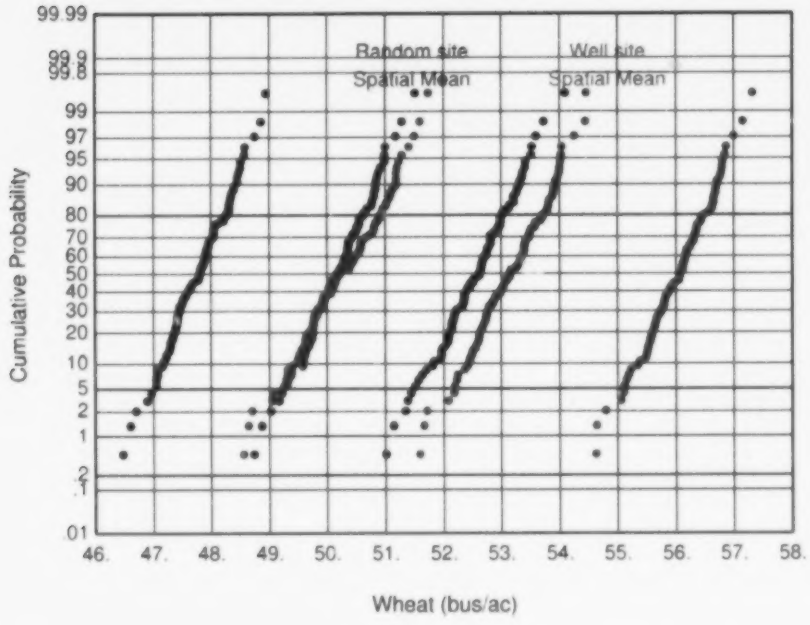
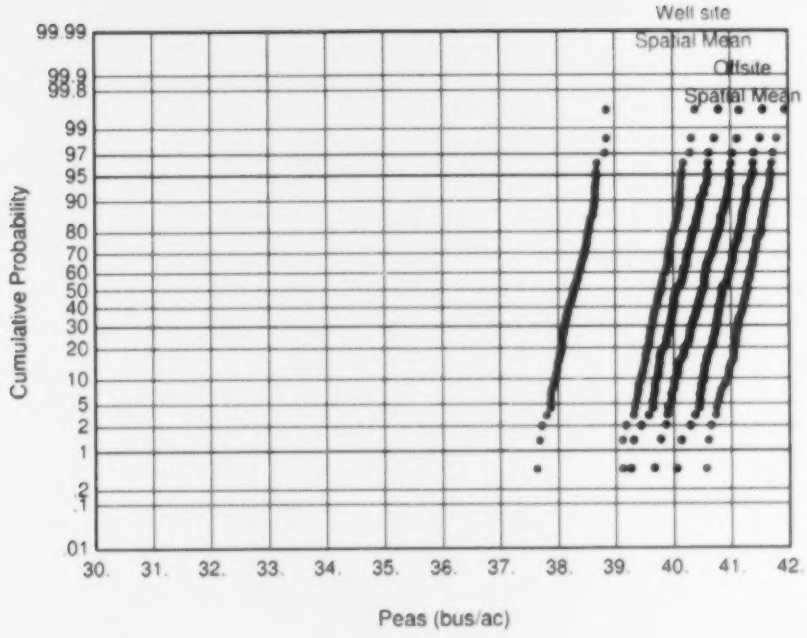


Figure A23 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 1 in 2000 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



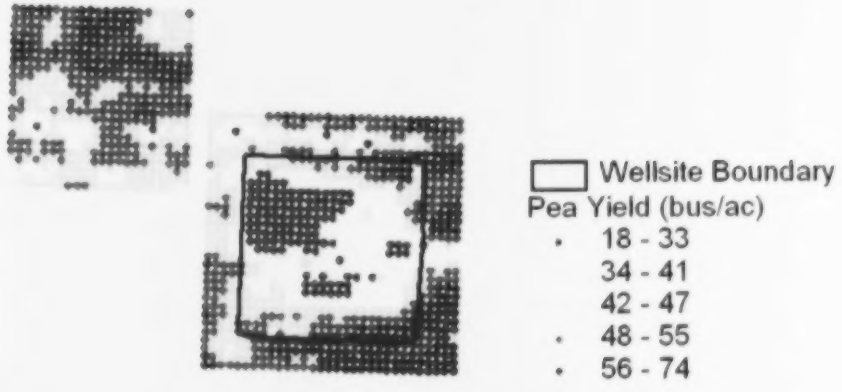


Figure A24 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 1 in 2003 significant at the 95% level for 101 models comparing the well site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Pea yield map from field 1 for the well site, offsite and random site (bottom figure).

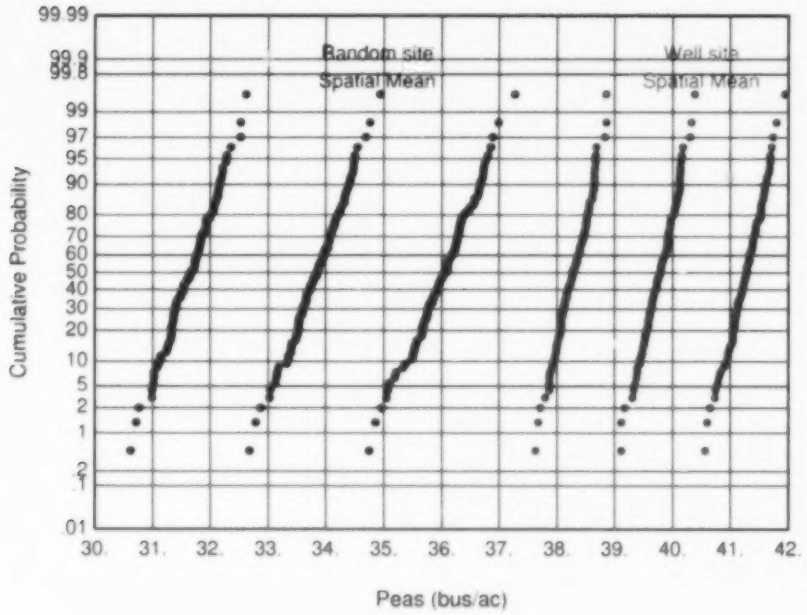
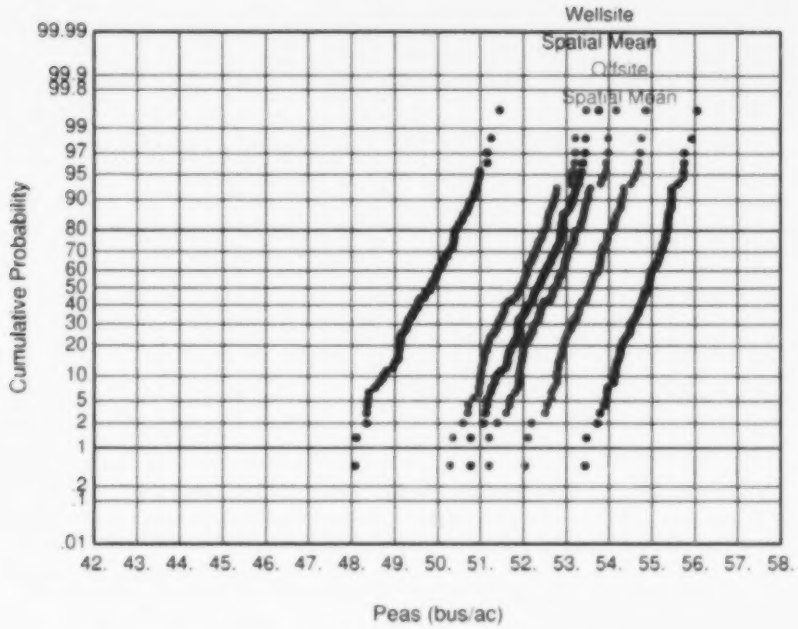


Figure A25 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 1 in 2003 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



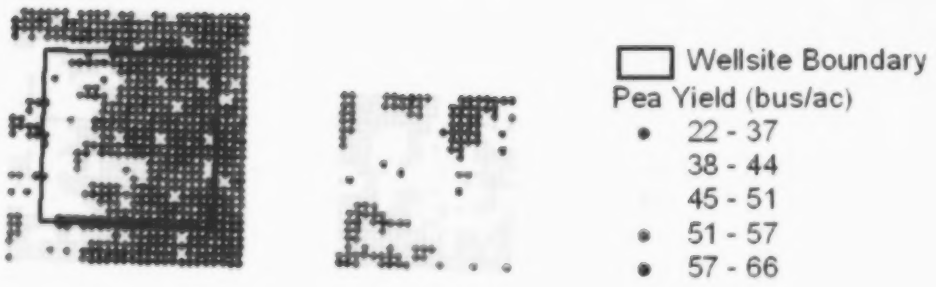


Figure A26 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 2 in 2005 significant at the 95% level for 101 models comparing the well site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Pea yield map from field 2 for the well site, offsite and random site (bottom figure).

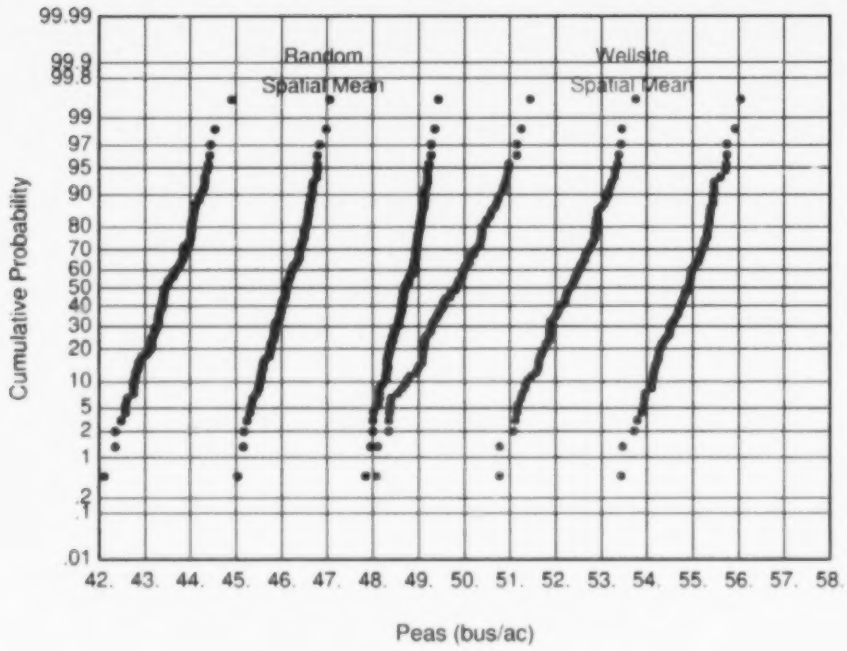
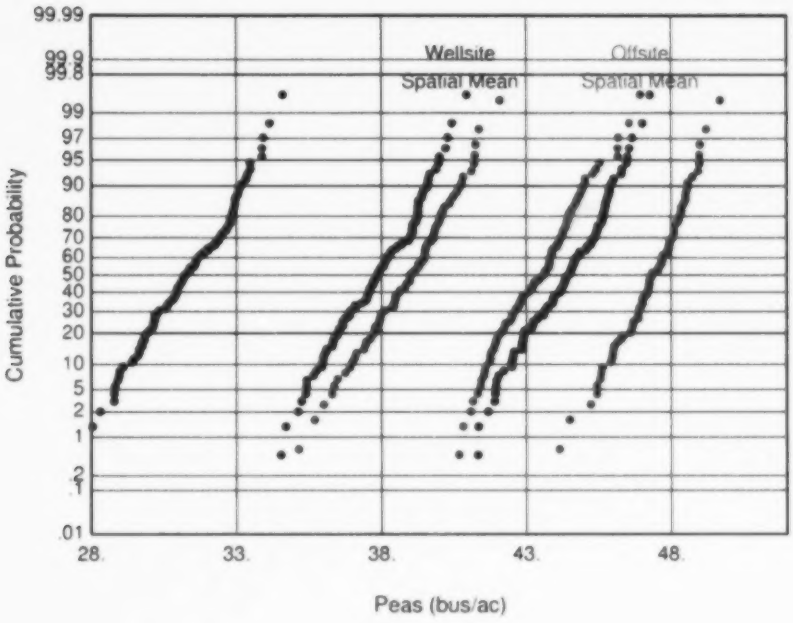


Figure A27 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 2 in 2005 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



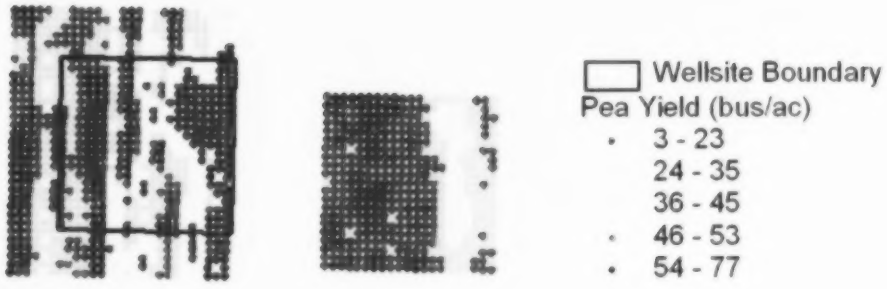


Figure A28 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 3 in 2000 significant at the 95% level for 101 models comparing the well site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Pea yield map from field 3 for the well site, offsite and random site (bottom figure).

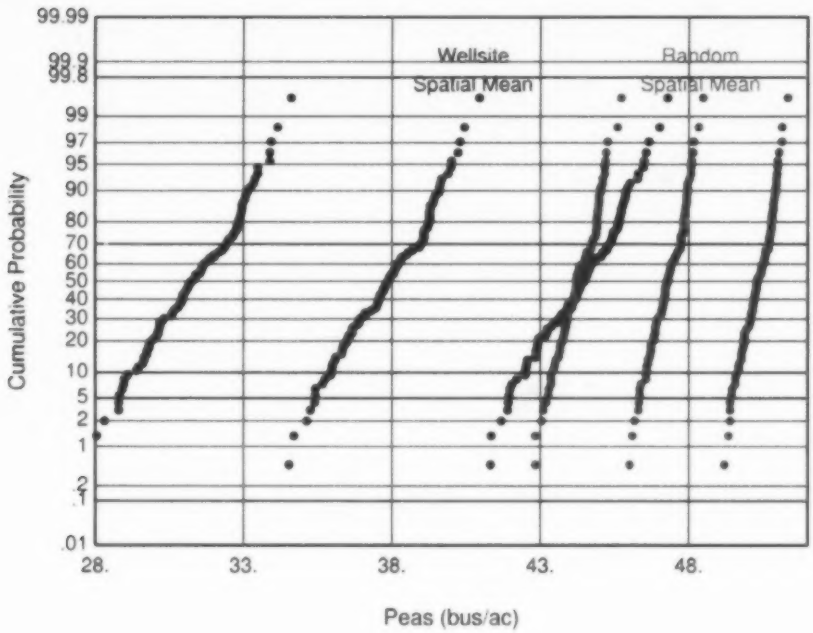
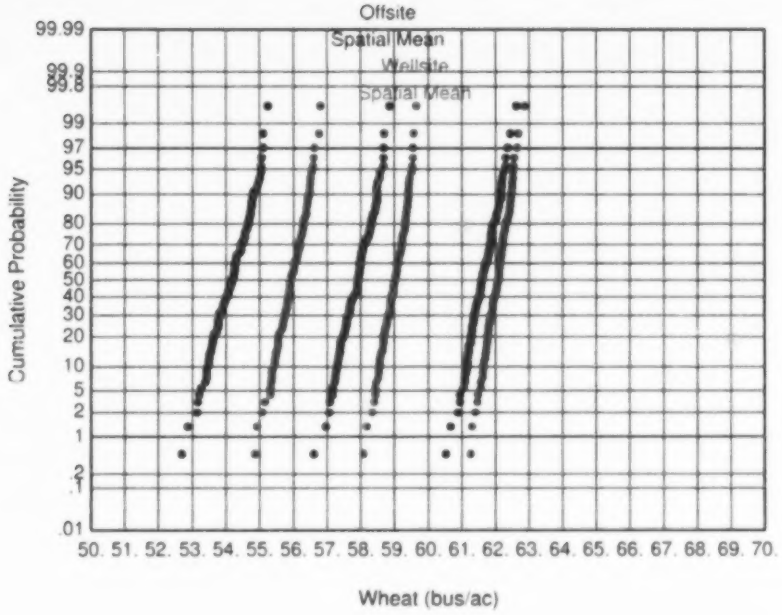


Figure A29 Confidence intervals and spatial averages for pea yield (bushels/acre) from field 3 in 2000 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



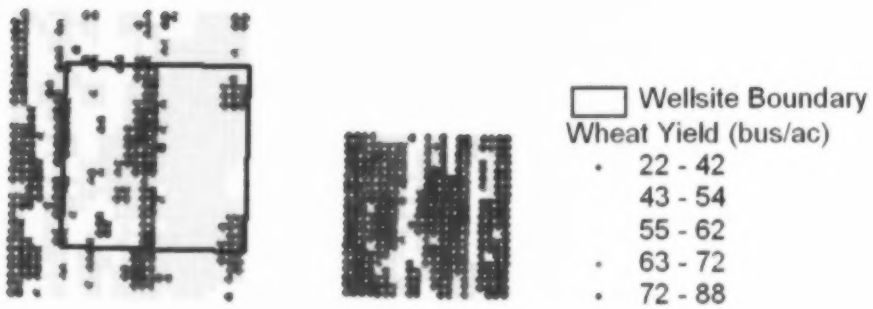


Figure A30 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 3 in 2004 significant at the 95% level for 101 models comparing the well site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Wheat yield map from field 3 for the well site, offsite and random site (bottom figure).

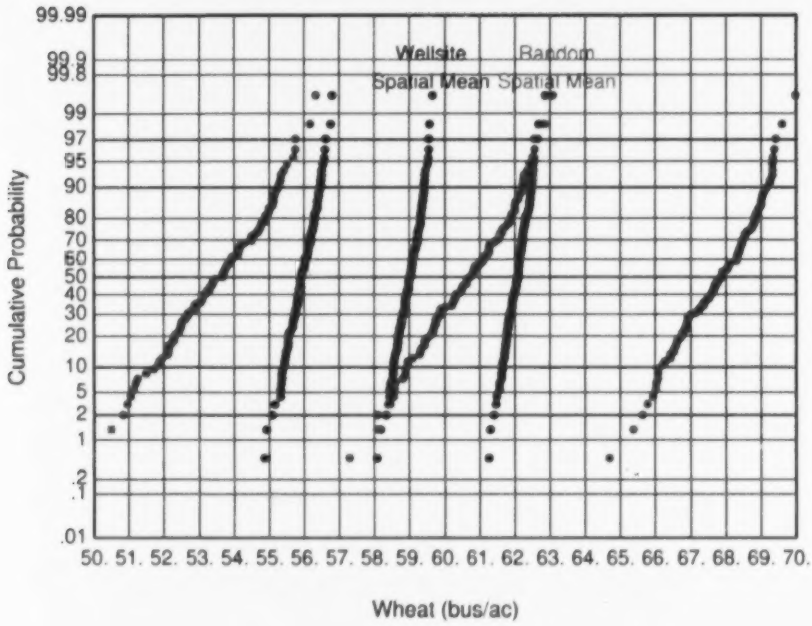
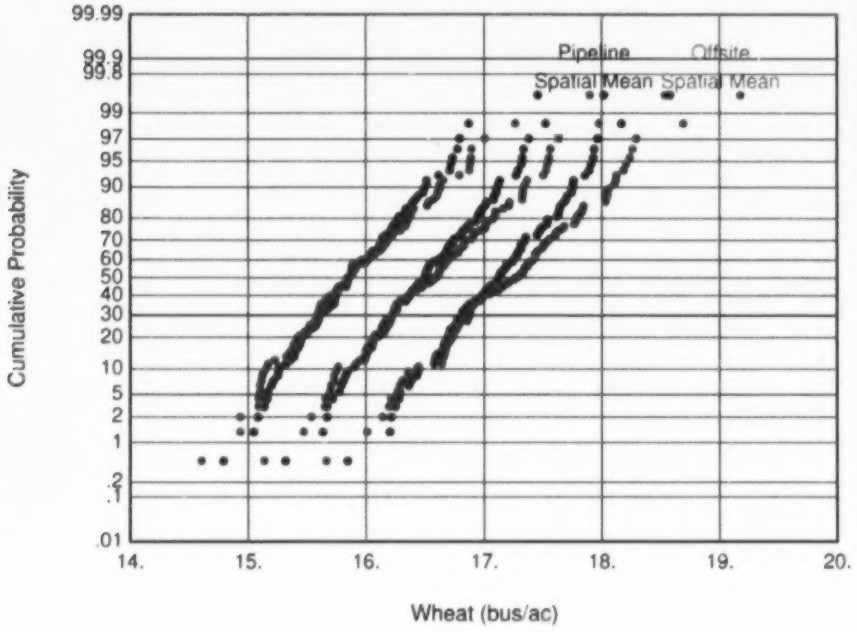


Figure A31 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 3 in 2004 significant at the 95% level for 101 models comparing the well site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



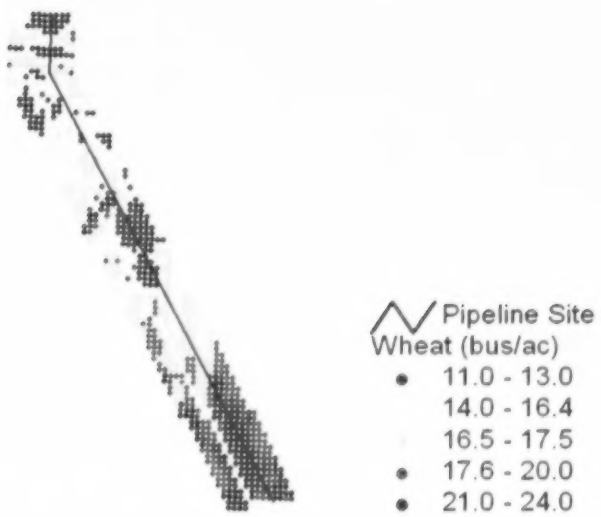


Figure A32 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 4 in 2002 significant at the 95% level for 101 models comparing the pipeline to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Wheat yield map from field 4 for the well site, offsite and random site (bottom figure).

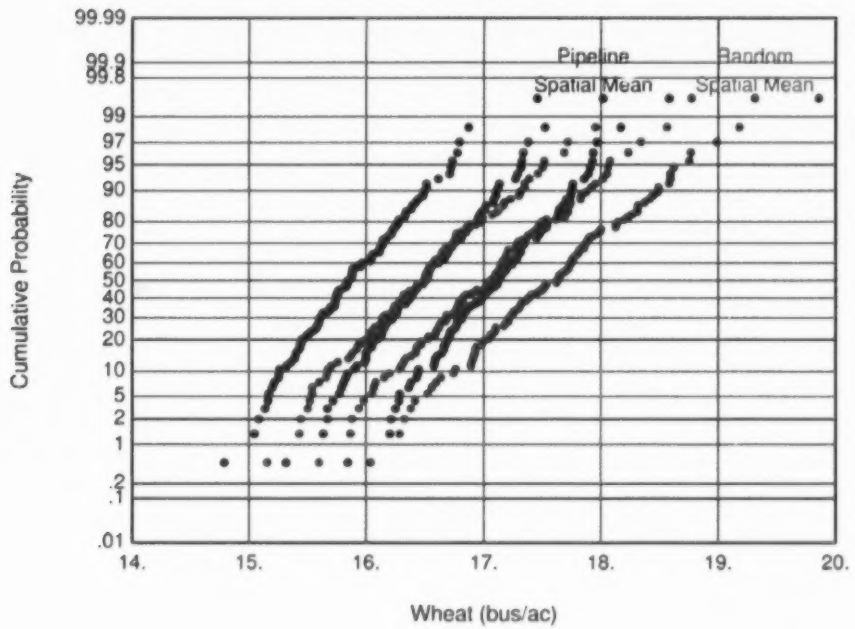
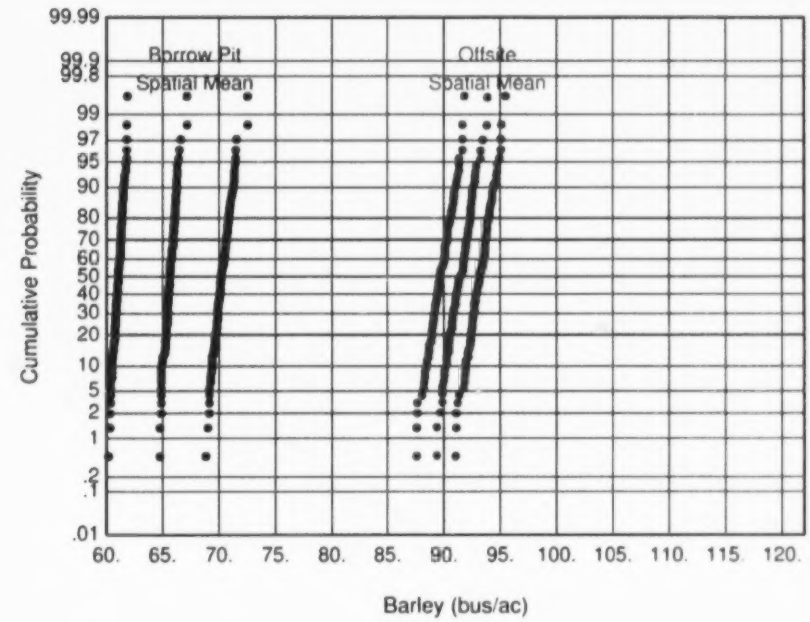


Figure A33 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 4 in 2002 significant at the 95% level for 101 models comparing the pipeline to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



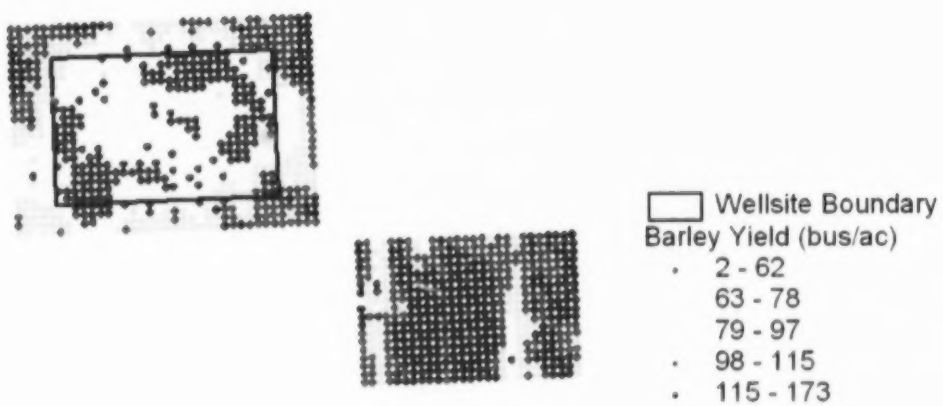


Figure A34 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 5 in 2003 significant at the 95% level for 101 models comparing the borrow pit site to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Barley yield map from field 5 for the well site, offsite and random site (bottom figure).

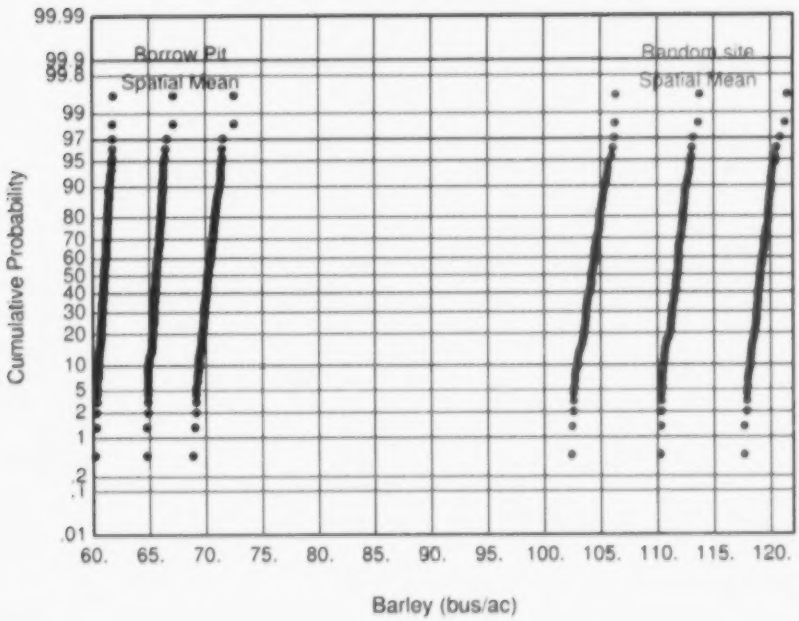
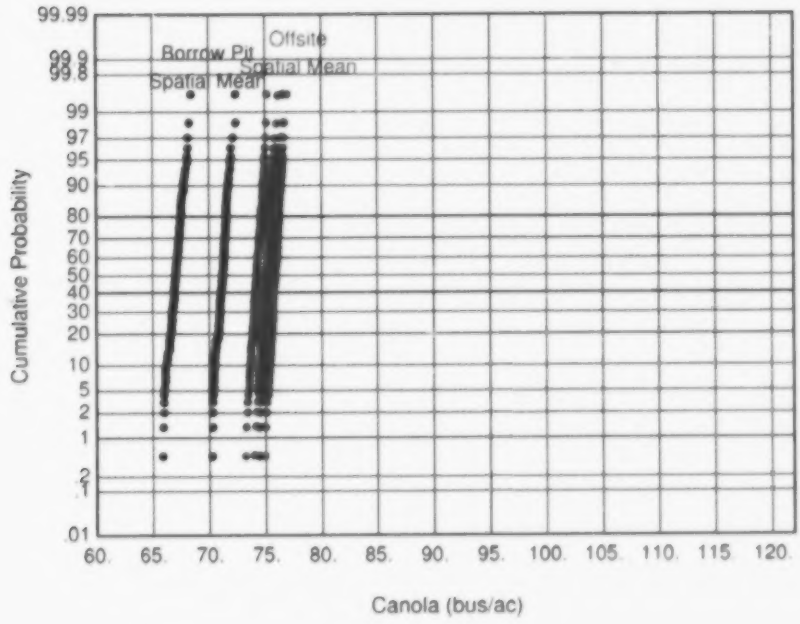


Figure A35 Confidence intervals and spatial averages for barley yield (bushels/acre) from field 5 in 2003 significant at the 95% level for 101 models comparing the borrow pit site to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



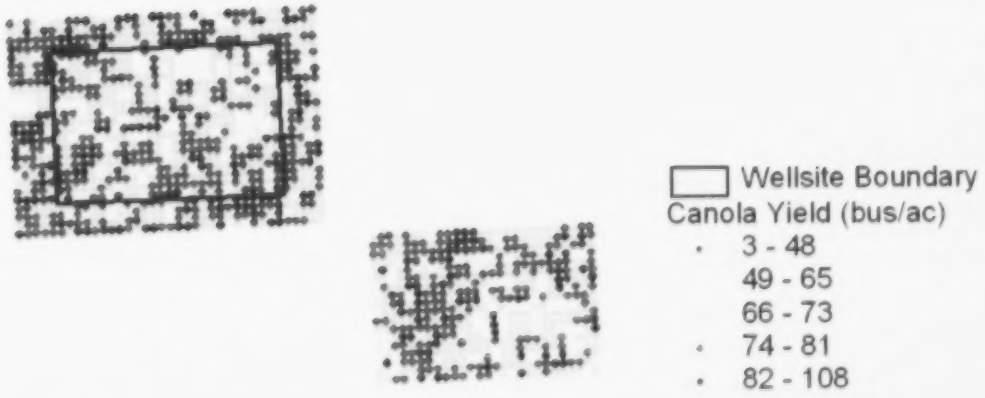


Figure A36 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 5 in 2005 significant at the 95% level for 101 models comparing the borrow pit to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Canola yield map from field 5 for the well site, offsite and random site (bottom figure).

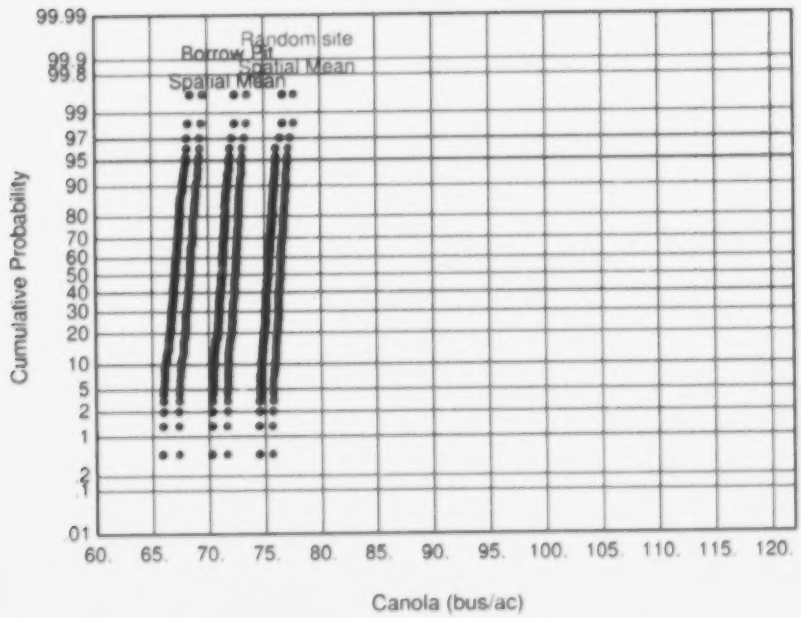


Figure A37 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 5 in 2005 significant at the 95% level for 101 models comparing the borrow pit to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.

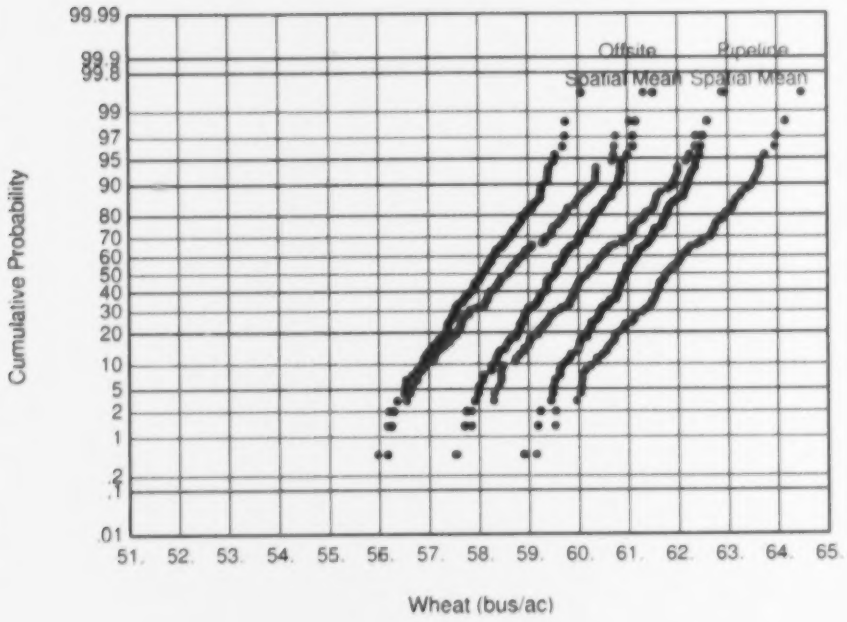


Figure A38 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 6 in 2005 significant at the 95% level for 101 models comparing the pipeline to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Wheat yield map from field 6 for the well site, offsite and random site (bottom figure).

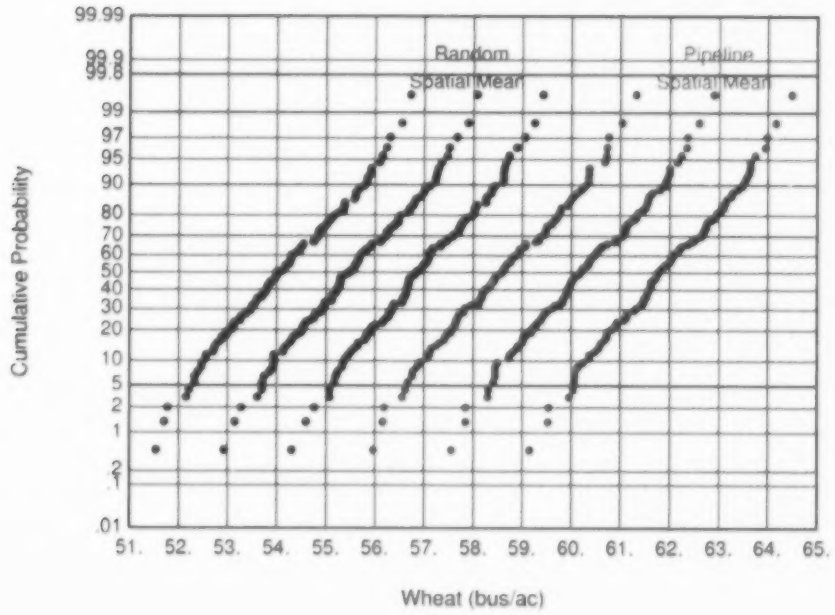


Figure A39 Confidence intervals and spatial averages for wheat yield (bushels/acre) from field 6 in 2005 significant at the 95% level for 101 models comparing the pipeline to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.

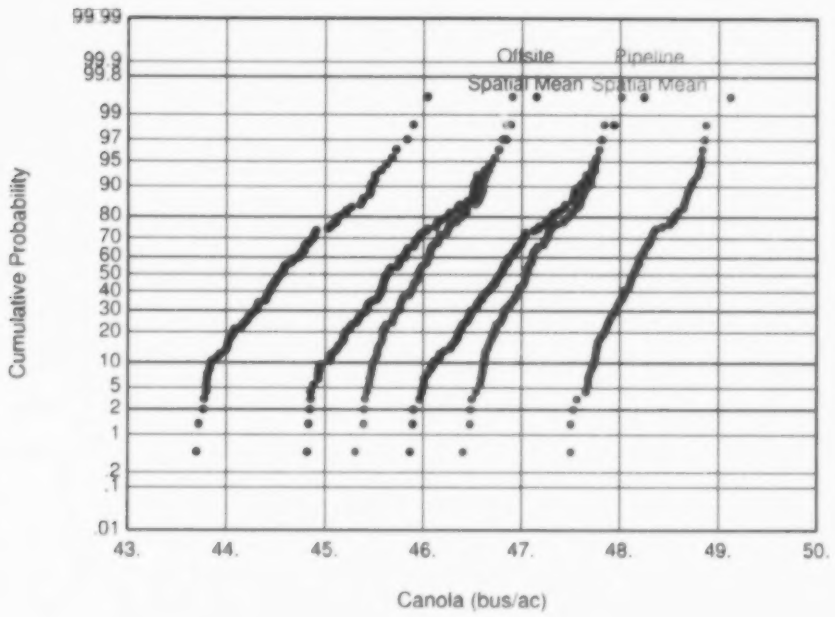


Figure A40 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 7 in 2004 significant at the 95% level for 101 models comparing the pipeline to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Canola yield map from field 7 for the well site, offsite and random site (bottom figure).

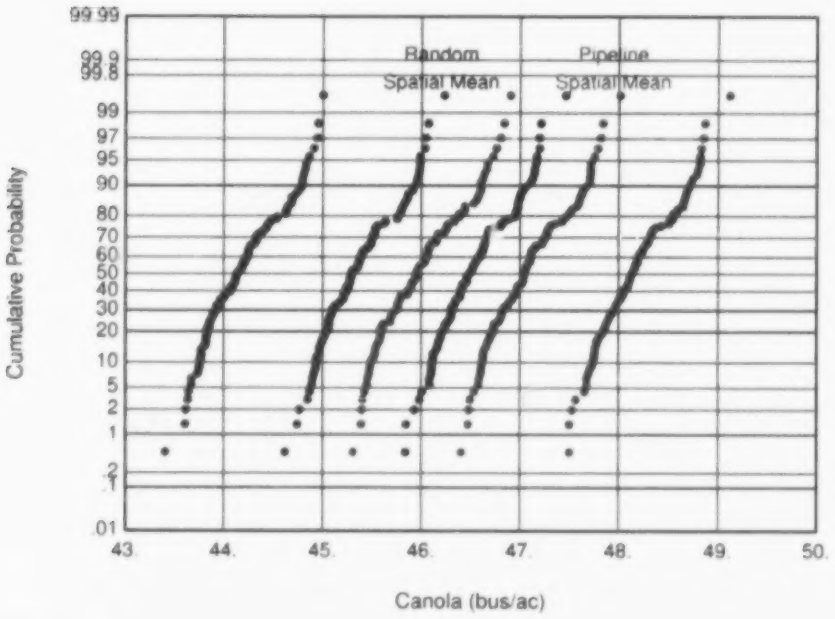
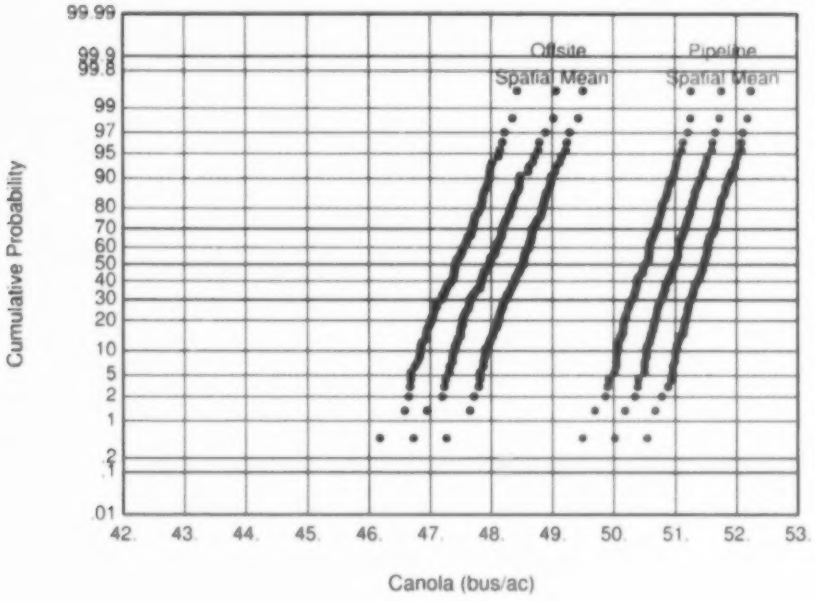


Figure A41 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 7 in 2004 significant at the 95% level for 101 models comparing the pipeline to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.



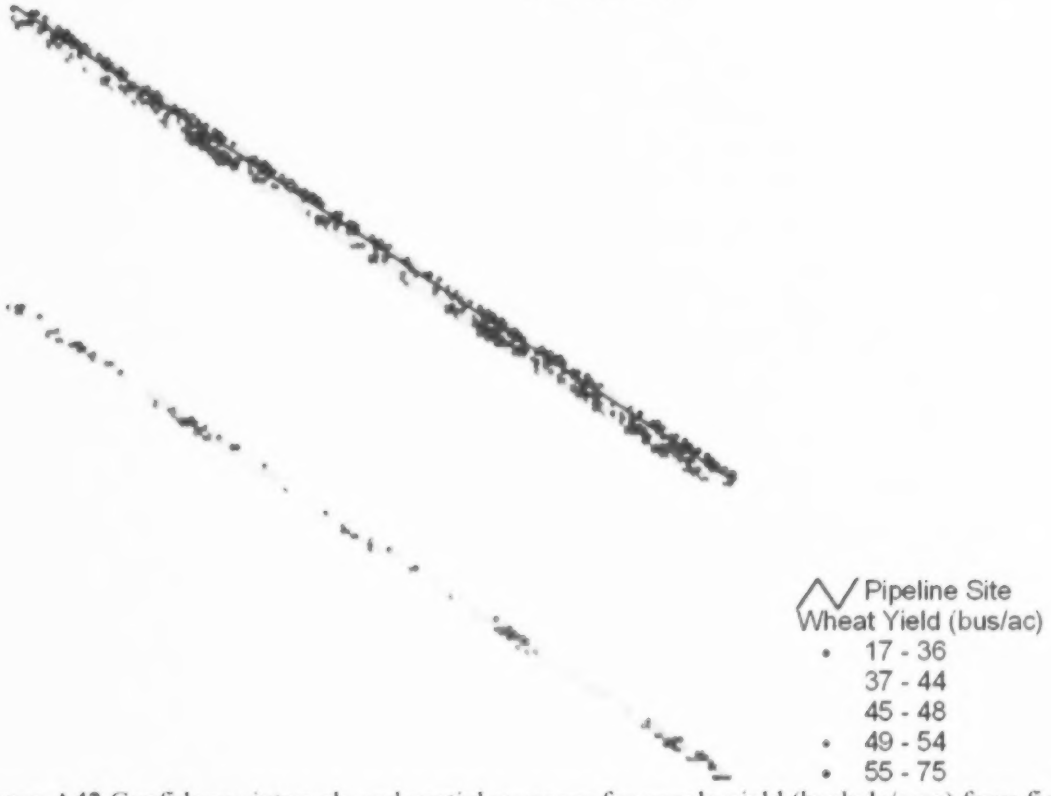


Figure A42 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 8 in 2004 significant at the 95% level for 101 models comparing the pipeline to the offsite site area (top figure). The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals. Wheat yield map from field 8 for the well site, offsite and random site (bottom figure).

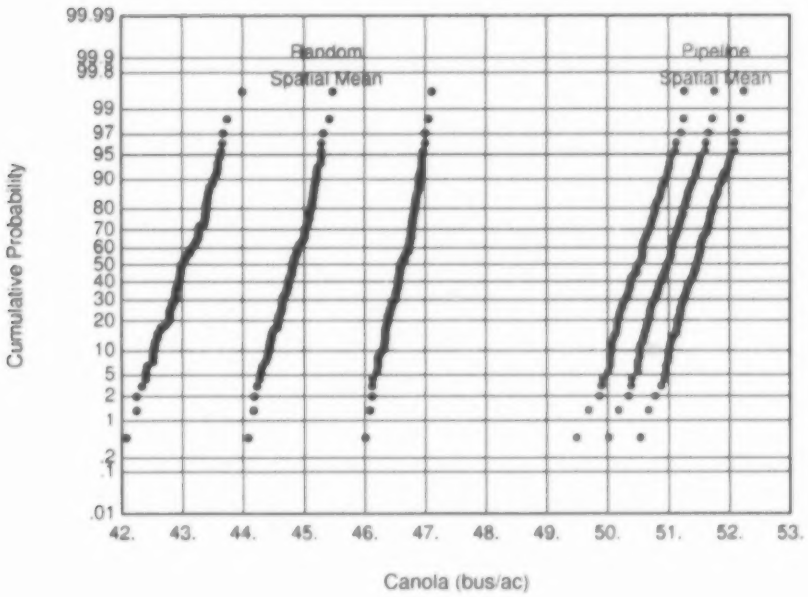


Figure A43 Confidence intervals and spatial averages for canola yield (bushels/acre) from field 8 in 2004 significant at the 95% level for 101 models comparing the pipeline to the random site area. The center set of dots is the spatial average while the dots on either side of the spatial average represent confidence intervals.